TITLE

Uncovering an allosteric mode of action for a selective inhibitor of human Bloom syndrome protein

SHORT TITLE

An allosteric inhibitor of human BLM

AUTHORS

X. Chen[1], Y. Ali[2], C.E.L. Fisher[1], R. Arribas-Bosacoma[1], M.B. Rajasekaran[2], G. Williams[2], S. Walker[2], J.R. Booth[2], J.J.R. Hudson[2], S.M. Roe[3], L. H. Pearl[1], S. E. Ward[2,4*], F. M. G. Pearl[3*], A.W. Oliver[1*]

AFFILIATIONS

1. Cancer Research UK DNA Repair Enzymes Group, Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, BN1 9RQ, United Kingdom.
2. Sussex Drug Discovery Centre, School of Life Sciences, University of Sussex, Falmer, BN1 9QJ, United Kingdom.
3. School of Life Sciences, University of Sussex, Falmer, BN1 9QG, United Kingdom.
4. Medicines Discovery Institute, Park Place, Cardiff University, Cardiff, CF10 3AT, United Kingdom.

* wards10@cardiff.ac.uk, f.pearl@sussex.ac.uk or antony.oliver@sussex.ac.uk

ABSTRACT

BLM (Bloom syndrome protein) is a RECQ-family helicase involved in the dissolution of complex DNA structures and repair intermediates. Synthetic lethality analysis implicates BLM as a promising target in a range of cancers with defects in the DNA damage response, however selective small molecule inhibitors of defined mechanism are currently lacking. Here we identify and characterise a specific inhibitor of BLM's ATPase-coupled DNA helicase activity, by allosteric trapping of a DNA-bound translocation intermediate. Crystallographic structures of BLM-DNA-ADP-inhibitor complexes identify a hitherto unknown interdomain interface, whose opening and closing are integral to translocation of ssDNA, and which provides a highly selective pocket for drug discovery. Comparison with structures of other RECQ helicases provides a model for branch migration of Holliday junctions by BLM.

MAIN TEXT

INTRODUCTION

RECQ helicases catalyse the unwinding of duplex DNA with 3' to 5' directionality, driven by energy liberated by ATP-hydrolysis. As well as simple DNA duplexes, the various members of the RECQ helicase family (BLM, WRN, RECQ1, RECQ4 and RECQ5 in humans) are able to unwind DNA within a range of complex DNA structures and DNA repair intermediates, including: forks, bubbles, triple helices, displacement (D)-loops, G-quadraplexes, and 3 or 4-way Holliday junctions (extensively reviewed in[1,2]).

RECQ-helicases are strongly implicated in the maintenance of genomic integrity, principally through their participation in the homologous recombination (HR) pathway for repair of DNA double-strand breaks and restart of collapsed or blocked replication forks (reviewed in[1,3]), but also have roles in toleration of microsatellite instability[4,5] and sister chromatid decatenation[6]. Defects in RECQ-family members are responsible for rare genetic diseases displaying substantial genomic instability and cancer predisposition[7]. Loss of function of WRN underlies the complex progeria Werner Syndrome; defects in BLM underlie Bloom Syndrome, which is characterised by growth retardation and immunodeficiency; and defects in RECQ4 are associated with Rothmund-Thompson syndrome, which displays growth retardation, skeletal abnormalities and premature ageing.

A number of experimental and computational studies have implicated RECQ helicases – primarily BLM and WRN - as potential targets for cancer therapy, due to the synthetic lethality of their silencing or downregulation with genetic defects inherent in a range of different cancers [4,5,8-12]. Despite the therapeutic opportunities this presents, no drugs targeting RECQ helicases have yet been licensed, although potential leads have been reported[13-15].

Here we determine the mode of action for two reported inhibitors of BLM – ML216[13,14] and a substituted benzamide (compound **2**). While ML216 appears to act, at least in part, through direct DNA binding and has poor specificity, we find that **2** and derivatives thereof are highly specific binders of a defined BLM-DNA complex. Crystallographic analysis of the BLM-DNA-**2** complex identifies a novel allosteric binding site and reveals a distinctive conformational step in the helicase mechanism, that can be trapped by small-molecules. These data pave the way for the development of allosteric inhibitors of BLM helicase with the potential to generate trapped and highly cytotoxic BLM-DNA complexes.

RESULTS

Compound identification and screening

A series of compounds that targeted the helicase activity of human BLM were identified in a quantitative high-throughput screen (qHTS)[14], where the results were made publicly available from the PubChem data repository [https://pubchem.ncbi.nlm.nih.gov/bioassay/2528].

Filtering the 627 reported active compounds for preferential physicochemical properties (e.g. Lipinski's rule of five) and excluding those with potential pan-assay interference activity (PAINS) allowed us to group the compounds into several distinct clusters according to chemical similarity. The inhibitory activity of exemplars from each cluster were tested in a fluorescence-based DNA unwinding assay[14] against recombinant human BLM-HD (HD = helicase domain; amino acids 636-1298). However, only a single compound produced an IC_{50} lower than 10 μM (compound **1**, IC_{50} = 4.0 μM; **Fig.1A**).

We synthesised and purified 6 close analogues of this compound with the aim of generating preliminary structure-activity relationship data and confirmed their inhibitory activity in the unwinding assay (**Materials and Methods).** Compounds **2** to **6** inhibited the 3' → 5' helicase activity of recombinant human BLM-HD with IC_{50} values ranging from 2.2 to ~60 μM, whereas **7** did not inhibit BLM-HD over the concentration range tested (**Fig. 1A and S1, Table 1)**. An IC_{50} of 4 μM was determined for ML216, a compound reported to be a semi-selective inhibitor of human BLM[13,14], which was included as a positive control (**Fig. 1A)**.

In a malachite green-based assay that measures ATP turnover, we observed robust stimulation of hydrolysis by BLM-HD when the protein was incubated with a short single-stranded 20-base oligonucleotide (**Fig. 1B**). Here, we determined IC_{50} values ranging from 3.2 to ~50 μM for each of our active analogues and 4.4 μM for ML216 (**Fig. 1E and S1, Table 1**). Whilst the values of IC_{50} obtained in our orthogonal assay did not agree in absolute value with those determined in the first, it ranked each analogue with a similar order of potency.

Biophysical analysis of compound binding

We could readily observe changes in fluorescence, indicative of binding, upon titration of both ADP and ATP-γS into BLM-HD using microscale thermophoresis (MST, **Fig. S2A**). We could not however observe any interaction for our most potent compound **2**.

In the absence of biophysical evidence for binding, we sought to confirm that **2** wasn't just a false positive generated by interference with the fluorescent readout of the unwinding assay. An alternative gel-based assay allowed direct visualisation of the conversion of a forked DNA-

duplex into its component single-stranded oligonucleotides via the helicase activity of BLM-HD (**Fig. 2A**). Titration of **2** clearly inhibited production of the single-stranded DNA product in a dose-dependent manner, with a calculated IC_{50} of 1.8 μM (**Fig. 2B**).

Mode of Inhibition: DNA-interaction

Another potential false positive could be generated by compounds that bind directly to DNA, thus preventing BLM-HD from productively engaging with its substrate. To test this hypothesis, we used a commercial DNA-unwinding assay that utilises recombinant Topoisomerase I (Topo I) to relax a supercoiled plasmid. Compounds that intercalate or bind to the major or minor groove of the plasmid DNA prevent relaxation.

At the manufacturer's recommended concentration of 200 μM, the positive control m-Amsacrine (mAMSA) strongly inhibited relaxation of the supercoiled plasmid. In contrast, no effect was observed with **2** at the same concentration. However, partial inhibition of relaxation could be observed for a reaction containing ML216 (**Fig. 2C**). To confirm this observation, we purchased ML216 from an alternative commercial supplier (ML216-A) and also resynthesized and purified the compound in-house (ML216-B; **Materials and Methods**). In both cases, a similar level of inhibition was observed when the compounds were included in the relaxation assay, indicating that that this was both a real and reproducible effect (**Fig. 2C**).

To explore further the possibility that ML216 might interact directly with DNA, we tested its ability to displace SYBR Green II (SG2) from a DNA substrate in a dye displacement assay[16,17]. When SG2 binds to DNA, a concomitant increase in its fluorescence can be measured. If an added compound can compete with the dye for binding to the DNA, a corresponding decrease in the fluorescent signal is observed. We titrated ML216 into a forked-50mer dsDNA substrate, that had been pre-incubated with SG2, observing a clear time- and dose-dependent displacement of the dye, indicating that ML216 can directly interact with a DNA substrate (**Fig. 2D**).

Mode of Inhibition: ATP-competitive

With confidence that **2** was, in fact, a *bona fide* inhibitor of BLM, we repeated the unwinding assay in the presence a 10-fold higher concentration of ATP to examine if the compound was directly competitive with nucleotide binding to the active site of the enzyme. As the resulting IC_{50} value was identical to that previously determined, it ruled out this mode of inhibition, and suggested that the compound bound elsewhere **(Fig. S2B**).

121 **Mode of Inhibition: Non-competitive**

122 ATP-turnover experiments, under Michealis-Menten conditions, allowed us to generate a
123 Lineweaver-Burk plot with data taken from DNA substrate titrations in the presence of 0, 5 and
124 10 μM of **2**. The resultant plot indicated a non-competitive (allosteric) mode of inhibition for **2**
125 (**Fig. 2E**). With this information, we postulated that **2** might only bind to BLM-HD when it was
126 engaged with a DNA substrate. We therefore revisited MST, first confirming the interaction of
127 BLM-HD with single-stranded DNA oligonucleotides 15 and 20 bases in length (**Fig. 2F**). We
128 next titrated **2** into the two pre-formed BLM-HD/ssDNA complexes. This time changes in
129 fluorescent signal could be detected, confirming our hypothesis, with dissociation constants
130 of 1.7 and 2.6 μM determined for the interaction with the 15mer and 20mer respectively (**Fig.**
131 **2G**).

132

133 **Mode of Binding: *BLM-HD$^{\Delta WHD}$***

134 We created the expression construct BLM-HD$^{\Delta WHD}$ to remove the conformationally flexible
135 Winged Helix domain (WH) that requires the presence of either a stabilising nanobody, or
136 interaction with a large DNA substrate to facilitate crystallogenesis[18,19] replacing it with a short
137 poly-(glycine/serine) linker that serves to connect the Zinc-binding domain (Zinc) directly to
138 the Helicase and RNAse C-terminal domain (HRDC, **Fig. 3A**). In validation of this approach,
139 we were able to crystallise the protein in complex with ADP and magnesium co-factor, and to
140 determine its structure at a resolution of 1.53 Å; a significant increase in resolution over
141 structures previously deposited in the PDB (4CDG, 2.8 Å; 4CGZ, 3.2 Å; 4O3M, 2.3 Å; see
142 **Table S1**).

143

144 Superposition of the structures of BLM-HD (PDB: 4CDG) and BLM-HD$^{\Delta WHD}$ produces a rmsd
145 of 0.86 Å over 2450 atom positions (D1 + D2 + Zn; PyMOL), indicating the overall conformation
146 and geometry of the two recombinant proteins is highly similar, despite deletion of the WH
147 domain (**Fig. S3A**). Furthermore, BLM-HD$^{\Delta WHD}$ binds both ssDNA-15mer and **2** with a similar
148 affinity to that of BLM-HD (**Fig. S3B**).

149

150 ***Crystal structure of BLM-HD$^{\Delta WHD}$ in complex with compound 2 and ssDNA***

151 We crystallised **2** in complex with BLM-HD$^{\Delta WHD}$, ADP/magnesium co-factor, and ssDNA-
152 15mer (liganded complex); determining its structure at a resolution of 3.0 Å (**Table S1).** The
153 complex crystallised in space group P1, with 6 molecules of BLM-HD$^{\Delta WHD}$ and associated
154 ligands forming the asymmetric unit. Interestingly, the co-crystallised ssDNA-15mer helped
155 drive formation of the crystal lattice, due to its partial self-complementary at the 5' end (5′-
156 CGTAC-3′) that serves to form four consecutive base pairs between two oligonucleotides

157 (**Fig. 3B**); the cytosine at the 5' end of the oligonucleotide is not readily discernible in electron
158 density maps and is therefore likely to be disordered. An extensive series of interactions are
159 made to the bound nucleic acid by amino acids from all four sub-domains of the BLM-HD$^{\Delta WHD}$
160 expression construct (**Fig. S3C**).
161
162 Compound **2** sits in a small pocket found on the opposite face of the protein to that which
163 binds nucleotide (**Fig. 3B, 3B inset)**, and integrates amino acid side chains from both the D1
164 and D2 subdomains of the helicase core, as well as several from the Zn-binding domain. The
165 oxygen of the amino group at the centre of the compound is hydrogen-bonded to the side
166 chain of Asn1022, whilst the nitrogen of the same moiety is in hydrogen-bonding distance to
167 both the backbone oxygen and side chain of Ser801. The side chains of His805 and Thr1018
168 stack up against, and provide Van der Waals contacts to, the central ring system of the 3-
169 amino-4,5-dimethylbenzenesulfonamide pendant group as part of a pocket lined by the side
170 chains of residues Asp806, His1014, Thr1015, His1019 **(Fig.3C**). The nitrogen of the
171 sulphonamide group is hydrogen-bonded to the side chain of Asp840, which itself is bonded
172 to the side chain of His805. The 2-methyl-thiazole moiety of **2** sits against the surface of the
173 alpha-helix containing Gly972 and is sandwiched by additional packing interactions with the
174 side chains of Gln802 and Glu971. The side chains Thr832 and His798 also contribute to this
175 section of the binding pocket, which is 'capped' by Gln975. The central benzene ring of **2** is
176 also contacted by the side chains of Gln802 and Glu971.
177
178 ***Reconfiguration of the Aromatic-Rich Loop***
179 The aromatic-rich loop (ARL) is a highly conserved motif in RecQ helicases that serves as a
180 molecular 'sensor', detecting binding of single-stranded DNA and coupling it to structural
181 rearrangements that enable ATP hydrolysis[20,21]. In our high-resolution structure of BLM-
182 HD$^{\Delta WHD}$, the ARL is disordered and is not visible in electron-density maps (**Fig. 4A**). By
183 contrast, it can be fully modelled in the liganded complex, but its conformation is distinct from
184 that observed in PDB entries 4CDG and 4CGZ where the single-stranded extension of bound
185 nucleic acid substrates does not extend across to the D1 domain (**Fig. 4B and S4A**).
186
187 Structures with a high degree of structural similarity to the liganded complex were identified
188 with PDBeFold[22]. The search produced PDB entries 6CRM and 4TMU, with Q-scores of 0.47
189 and 0.46 respectively, which both describe structures of the catalytic core of *Cronobacter*
190 *sakazakii* RecQ helicase (CsRecQ) in complex with different DNA substrates[20,23].
191 Comparison, in each case, reveals a close to identical conformation of the ARL to that
192 observed in our liganded complex, as well as nucleic acid interactions that include the D1
193 domain (**Fig. S4B and C**).

194 For CsRecQ, Manthei *et al.* described concerted movements of residues Phe158 and Arg159,
195 within the ARL to interact with the 3' single-stranded extension of their co-crystallised DNA
196 substrate[20]; in our liganded complex the equivalent residues undergo a similar transition
197 (Phe807 and Arg808 respectively). We observe that Phe158 moves to make base-stacking
198 interactions with G8 and A9 of the bound ssDNA-15mer. By comparison to 4CDG, we also
199 see that Arg808 switches from interacting with Asp806 of the ARL to the backbone oxygen of
200 Pro715 and the side chain of Glu768 (**Fig. 4A and 4B**), Notably, mutation of residues
201 equivalent to Arg808 or Glu768 in EcRecQ (Arg159 and Glu124) have been shown to perturb
202 enzyme function[20,21]. Interestingly, in BLM, Asp806 is 'freed' to interact with the side chain of
203 His1029, a residue within the Zn-binding domain (**Fig. 4B**).
204
205 In their 2017 paper, Newman *et al.*, proposed a general mechanism for the helicase activity
206 of human RECQ5 (hsRECQ5) based on their analyses of the available crystal structures for
207 RecQ-family helicases; indicating that 4TMU best represents a 'pre-ATP hydrolysis'
208 conformational state[24]. In their structure of unliganded (apo) hsRECQ5 they report a polar
209 contact between Gln345 and His160 of the ARL, which is subsequently broken as a result of
210 binding to ATP, serving to free the ARL to interact productively with the single-stranded 3'-
211 overhang of a bound DNA substrate. Importantly, they showed that mutation of glutamine 345
212 to alanine (Q345A) prevents stimulation of ATPase activity by binding to DNA, but does not
213 perturb the basal rate of hydrolysis[24]. The importance of this particular residue is perhaps
214 more apparent in *E. coli* RecQ, where mutation of the equivalent glutamine (Q322A)
215 essentially ablates all ATPase capability[20]. Notably, in our liganded structure, the equivalent
216 residues (Gln975 and His798 respectively) form part of the binding pocket for **2 (Fig. 3C, right**
217 **and 4A)**.
218
219 **HRDC from parked to DNA-engagement**
220 The HRDC (Helicase and RNase D C-terminal domain) was originally identified as a putative
221 nucleic-acid binding motif in both BLM and WRN (Werner syndrome helicase) and named in
222 part for its similarity to a domain found at the C-terminus of *E. coli* RNase D[25]. However, only
223 very weak ssDNA binding (K_d ~ 100 μM) has been reported for this domain in isolation[26].
224
225 In their paper describing the crystal structure of BLM in complex with DNA, Newman *et al*.
226 observe that the HDRC domain packs against a shallow cleft formed between the D1 and D2
227 domains of the helicase core, with the interface between the different modules being highly
228 polar in nature. Their follow-on small angle X-ray scattering experiments also indicate that the
229 HRDC domain of BLM is free to disassociate and re-bind to the helicase core[18]. In our high-
230 resolution crystal structure of BLM-HD$^{\Delta WHD}$ we observe the same 'parked' interaction for the

231 HRDC, even in the absence of the WHD (**Fig. S3A**), but in our liganded structure we see that
232 the HRDC domain swings across the face of the core helicase fold (**Fig. 5A**) to make a series
233 of polar interactions with the ssDNA-15mer, which is presented on the surface of the D1
234 domain of a second protomer within the asymmetric unit (**Fig. 3B and 5B**).
235
236 The side chains of HRDC residues Asn1242 and His1236, contained within the 'hydrophobic
237 3_{10} helix'[26], are hydrogen-bonded to the O4 group of the T12 base and the N3 of the G11 base,
238 respectively. The side chain of Phe1238 is also involved in an edge to ring-stacking interaction
239 with the G11 base. Each of these interactions is consistent with chemical shift changes
240 previously observed in HSQC spectra - as a result of titrating ssDNA into ^{15}N-labelled BLM-
241 HRDC[26] – suggesting that the observed interactions have biological relevance, and that our
242 structure represents the first to capture HRDC interactions with ssDNA. Furthermore, amino
243 acids residues Lys1227, Tyr1237, Thr1243, and Asn1239 are also in close proximity to the
244 bound DNA (**Fig. 5B**) and could be expected to undergo changes in chemical environment
245 upon interaction; again consistent with the reported perturbations in HSQC spectra[26]. Asn1239
246 might also be expected to pick up an additional contact with the 5'-phosphate of a subsequent
247 nucleotide in an extended substrate.
248
249 **Selectivity profile**
250 With a robust molecular understanding for the binding mode for **2**, we next examined if the
251 compound displayed selectivity for members of the RecQ-helicase family. In our ATP-turnover
252 assay we saw no inhibition of recombinant helicase domains (HD) from human WRN, human
253 RecQ5 or the unrelated *E. coli* helicase UvrD over the concentration range tested, whilst at
254 higher concentrations inhibitory effects started to appear against human RecQ1. In contrast,
255 ML216 robustly inhibited all four RecQ-family helicases tested and at higher concentrations
256 also affected UvrD; in line with, and in support of, our observation that ML216 is non-specific
257 and elicits at least part of its inhibitory effect by binding directly to DNA (**Fig. 6A**).
258
259 Whilst compound solubility prevented generation of a complete inhibition curve for UvrD and
260 thus a robust estimate of IC_{50}, the estimated Hill coefficient (*n*H) was close to 1 — in contrast
261 to those calculated for titrations of ML216 against the RecQ helicases, which were generally
262 steeper (ranging from 1.8 to 4.3), again suggesting that instead of a forming a 1:1 protein to
263 inhibitor complex, there is in fact, a more complex (possibly mixed) mode of binding for this
264 compound to this class of enzymes.
265
266
267

268 **Conformational trapping**

269 As binding of **2** has no direct effect on the ability of BLM-HD$^{\Delta WHD}$ to bind either ssDNA or
270 nucleotide, we hypothesised that it might act to 'lock' the helicase into a conformational state
271 where DNA substrates remain bound but cannot be unwound. In support of this idea, we
272 undertook MST assays with a labelled single-stranded oligonucleotide in the both the
273 presence and absence of **2**. Here, we observed a clear concentration-dependent decrease in
274 K_d in the FP experiment when **2** was added, consistent with a decrease in the off-rate for DNA-
275 binding, supporting our hypothesis that the compound acts to 'trap' BLM in its interaction with
276 ssDNA (**Fig. 6B**).

277

278 **DISCUSSION**

279 **Allosteric binding site**

280 Compound **2** binds to a small pocket — hereinafter referred to as the allosteric binding site
281 (ABS) — which is generated transitorily as part of the set of conformational changes that BLM
282 undergoes during its catalytic cycle; relative movements of the D1 and D2 domains with
283 respect to each other, upon interaction with nucleotide and DNA substrates, have previously
284 been described by Newman *et al*[24]. Only when the D1 domain of BLM interacts with single-
285 stranded DNA is the ARL in the correct conformation to permit compound binding, as then the
286 side chain of Trp803 is repositioned and no longer occludes / fills the ABS (**Fig. 4A, B**).

287

288 Gyimesi *et al*. determined that the rate-limiting step for the unwinding activity of BLM is a
289 structural transition between two ADP-bound states, which they suggest leads directly to
290 'stepping' along the single-stranded region of bound DNA substrates. They describe a so-
291 called 'open' state where ADP can freely dissociate from the enzyme, and a 'closed' state that
292 interconverts slowly to the open state in order to release ADP[27]. They also note that conversion
293 between these states is accelerated in the presence of ssDNA.

294

295 Taken as a whole, our experimental data would be consistent with the notion of **2** acting to
296 'trap' or stabilise the 'closed' state, preventing conversion to the 'open' state, ADP release and
297 thus DNA unwinding. Furthermore, compounds bound stably to the ABS should also sterically
298 prevent the ARL from reverting to its initial conformation / structurally disordered state found
299 at the beginning of the catalytic cycle (**Fig. 4A, B**).

300

301 However, this is perhaps at odds with the description of PDB entry 4TMU (CsRecQ, **Fig. S4**),
302 to which our liganded complex is the most similar, as an example of the pre-ATP hydrolysis
303 step of the RecQ-family helicase catalytic cycle[24]. A possible explanation, still consistent with
304 our demonstration of 'trapping', is that binding of **2** prevents the changes that normally result

305 as a consequence of ATP-hydrolysis and phosphate-release, keeping the enzyme in an
306 apparent pre-hydrolysis conformation. More detailed kinetic analyses will be required to
307 distinguish between these mode-of-action models.
308
309 **Selectivity through Helical Hairpin interactions**
310 Several different regions of BLM interact with **2** when it is bound to the ABS, including amino
311 acids from helicase motifs I and III, plus a short region just upstream of motif IV[28,29] (pre-motif
312 IV, **Fig. 7**). Unsurprisingly, the amino acid sequence identity of each region across the RecQ-
313 family is extremely high, and do not therefore provide a facile explanation for the observed
314 selectivity of **2** (**Fig. 6A**). In particular, two of the amino acid side chains involved in hydrogen
315 bonds with **2** are absolutely conserved in identity (motif I, Ser801; motif III, Asp840, **Figure**
316 **3C and 7**). Likewise, amino acids within these motifs involved in hydrophobic contacts with **2**
317 are also highly conserved in identity/chemical property. However, the third hydrogen bonding
318 interaction (made by Asn1022 to **2**) and its position within the 'Helical Hairpin' of the Zn-binding
319 domain provides some insight, as both the length and amino acid composition of this loop is
320 highly divergent across the RecQ-family and absent from RecQ4 (**Fig. 7**). There is no obvious
321 consensus for any of the residues, in equivalent positions to those in BLM, involved in
322 compound interaction. This observation provides a plausible route, though the addition or
323 alteration of chemical groups to the core scaffold of **2**, for generating potent and highly
324 selective inhibitors for the individual members of the RecQ-family of helicases.
325
326 **Speculative model for the involvement of the HRDC in unwinding DNA substrates**
327 During model building and evaluation of our liganded structure, we found that the nucleic-acid
328 interactions made by the different sub-domains of BLM-HD$^{\Delta WHD}$ explain how the HRDC might
329 contribute to the ability of BLM to unwind different types of DNA substrate — incorporating
330 information taken from our own structure, as well as that for the interaction of the WHD with
331 dsDNA from PDB entry 4CGZ (**Fig. 7B**).
332
333 Simple superposition of the two structures results in a clash of the HRDC (in our structure)
334 with the WHD, however this is readily resolved by a small horizontal translation of the HRDC
335 (roughly equivalent to adding an additional nucleotide to the single-stranded portion of the
336 bound DNA substrate). We also note that, in our liganded structure, the polarity of the single-
337 stranded DNA interacting with the HRDC is opposite to that of our model but with the
338 understanding that is actually dictated by the packing arrangement of the molecules that serve
339 to generate the crystal lattice; on examination, each of the observed HRDC interactions is fully
340 compatible with binding to ssDNA in either orientation.
341

342 However, using the trajectories for each of the bound DNA substrates as positional markers
343 allows generation of a model for unwinding of a simple DNA duplex. The double-stranded
344 portion of the substrate is held in place through the previously described set of interactions
345 with the WHD domain[18]. The β-hairpin of the WH serves to separate the DNA duplex, with
346 one strand 'actively' engaged with the D1/D2 domains of the helicase core and the Helical
347 Hairpin of the Zn-binding domain. The second 'inactive' strand passes along the opposite
348 face of the WH β-hairpin, to subsequently interact with the HRDC domain, potentially acting
349 to prevent reversion and re-annealing of the DNA duplex. The relatively poor binding affinity
350 of the HRDC for ssDNA is compatible with this model, as it would allow iterative release and
351 recapture of the 'inactive' strand as the DNA substrate is unwound.
352
353 A model for how BLM might unwind a Holliday Junction has previously been reported[30]; but
354 this treats the HRDC domain as a static object, leaving in it the 'parked' position and making
355 no interactions with nucleic acid. Data published subsequent to this paper has indicated that
356 the HDRC plays a more fundamental role, for example the charge-reversal mutation N1239D
357 to has been shown to ablate interaction of the HRDC with both ssDNA and a Holliday junction
358 substrate[26]. The HRDC has also been reported to confer DNA-structure specificity to BLM,
359 with Lys1270 playing a role in mediating interactions with DNA and for efficient dissolution of
360 double-Holliday junction substrates *in vitro* [31]. Consistent with this, our structural data reveals
361 that both Asn1239 and Lys1270 of the HDRC are poised to interact with a section of single-
362 stranded DNA just one to two nucleotides longer than that captured in our crystal lattice (**Fig.
363 5B**). Finally, our model also suggests the existence of a transient direct interaction between
364 the WHD and the HRDC.
365
366 Compound 2 itself, is unlikely to be suitable as a therapeutic agent, due to its poor ability to
367 penetrate the cell membrane of mammalian cells (data not shown). However, as the first
368 described *bona fide* selective allosteric inhibitor of human BLM, the understanding of its mode
369 of action will aid ongoing efforts to develop molecules targeting this class of enzymes for the
370 treatment of human disease.

371 **MATERIALS AND METHODS**

372 **Compound synthesis and purification**

373 Details for synthesis and purification of compounds is provided in supplementary information.

374

375 **Expression constructs**

376 Synthetic genes, codon-optimised for expression in *E. coli*, were purchased from GeneArt
377 [ThermoFisher Scientific, Loughborough, UK]. With the exception of RECQ5 (see below) the
378 coding sequence was subcloned into an in-house modified pET-17b vector at the NdeI and
379 EcoRI sites of the multiple cloning site.

380

381 *BLM-HD, BLM-HD$^{\Delta WHD}$:* pAWO-STREP-3C; a pET-17b expression vector modified to encode
382 an N-terminal human rhinovirus 3C-protease (HRV-3C) cleavable StrepII-affinity tag. BLM-HD
383 encodes amino acids 636-1298 of human BLM (UniProt ID: BLM_HUMAN), whereas BLM-
384 HD$^{\Delta WHD}$ encodes amino acids 636-1074 and 1231-1298.

385

386 *RECQ1-HD:* pAWO-His-TRX-3C*; a* pET-17b expression vector modified to encode an N-
387 terminal His$_6$-tagged *E. coli* Thioredoxin HRV-3C cleavable affinity/solubility tag. RECQ1-HD
388 encodes amino acids 49-616 of human RECQ1 (Uniprot ID: RECQ1_HUMAN).

389

390 *WRN-HD:* pAWO-His-SUMO-3C; a pET-17b expression vector modified to encode an N-
391 terminal His$_6$-tagged *S. cerevisiae* Smt3 (SUMO) HRV-3C cleavable affinity/solubility tag.
392 WRN-HD encodes amino acids 481-1521 of human WRN (Uniprot ID: WRN_HUMAN).

393

394 *RECQ5-HD:* The expression construct pNIC28-Bsa4-RecQL5 was obtained from the
395 Structural Genomics Consortium, Oxford [see https://www.thesgc.org/tep/RECQL5 for full
396 details]. RECQ5-HD encodes amino acids 11-526 of human RECQ5 (Uniprot ID:
397 RECQ5_HUMAN).

398

399 **Expression and purification**

400 *BLM-HD, BLM-HD$^{\Delta WHD}$*

401 *E. coli* strain BL21(DE3) [Promega, Southampton, UK] was transformed with the required
402 expression plasmid. A 'starter' culture was generated by inoculating a 250 ml glass
403 Erlenmeyer flask with 100 ml of Turbo-broth [Molecular Dimensions, Sheffield, UK]
404 supplemented with 50 μg/ml ampicillin. The culture was allowed to grow in an orbital-shaking
405 incubator set at 37°C, 220 rpm, until the absorbance at 600 nm reached 1.5. The culture was
406 then stored at 4°C until the following day. 12 ml of the 'starter' culture was used to inoculate

407 a 2 L Erlenmeyer containing 1 L of Turbo-broth supplemented with antibiotic as before. The
408 culture was grown until the absorbance at 600 nm reached 1.5, when the flask containing the
409 culture was placed on ice for a period of 30 minutes. During this time, the incubator
410 temperature was reduced to 20°C. After incubation on ice, isopropyl-β-D-thiogalactoside
411 (IPTG) was added to a final concentration of 0.4 mM, to induce protein expression. The flask
412 was then returned to the incubator, and the culture allowed to grow overnight at the reduced
413 temperature of 20°C. Cells were harvested by centrifugation after a period of 16 hours. The
414 resultant cell pellet was stored at -20°C until required.
415
416 The cell pellet arising from 2 L of culture was resuspended, on ice, in 25 ml of Buffer A (50
417 mM HEPES-NaOH, pH 7.5, 1 M NaCl, 0.5 mM TCEP, 0.5 mM EDTA) supplemented with a
418 protease inhibitor tablet [cOmplete EDTA-free Protease Inhibitor Cocktail Tablet; Roche,
419 Burgess Hill, UK]. The cells were lysed by sonication and insoluble material removed by
420 centrifugation. The resultant supernatant was applied to a 5ml Strep-Tactin Superflow Plus
421 Cartridge [Qiagen, Manchester, UK], pre-equilibrated with Buffer A. Unbound material was
422 application of 10 column volumes (CV) of Buffer A (50 ml). Retained proteins were then eluted
423 from the column by application of 5 CV of Buffer B (Buffer A supplemented with 5 mM
424 desthiobiotin). Fractions containing the required protein were identified by SDS-PAGE,
425 pooled, and then concentrated to a final volume of 2.5 ml using centrifugal concentrators
426 [Vivaspin 20, 5000 MWCO; Sartorius Stedim Biotech GmBH, Goettingen, Germany]. After
427 overnight cleavage of the affinity tag with human rhinovirus 3C-protease, the sample was
428 diluted to reduce the NaCl concentration to below 250 mM. This was applied to a 5 ml HiTrap
429 Heparin HP cartridge [GE Healthcare Life Sciences, Little Chalfont, UK], pre-equilibrated in
430 Buffer C [50 mM HEPES-NaOH, pH 7.5, 250 mM NaCl, 0.5 mM TCEP, 0.5 mM EDTA].
431 Unbound material was removed by washing the column with 10 CV of buffer C. A linear NaCl
432 gradient starting at a concentration of 250 mM and ending at 2000 mM, over 50 CV, was
433 applied to the column. Fractions containing the desired recombinant protein were identified,
434 pooled and concentrated as before. The concentrated sample was then applied to an HiLoad
435 26/600 Superdex 200 size exclusion chromatography column [GE Healthcare] pre-
436 equilibrated in Buffer D [20 mM HEPES-NaOH pH7.5, 250 mM NaCl, 0.5 mM TCEP]. Again,
437 fractions containing the desired recombinant protein were identified, pooled and concentrated,
438 then flash-frozen in aliquots in liquid nitrogen and stored at -80°C until required.
439
440 *RecQ1-HD, RecQ5-HD, WRN-HD*
441 Expression and purification of RecQ1-HD, RecQ5-HD and WRN-HD were achieved using
442 procedures similar to that used for BLM-HD, but with initial capture achieved using an IMAC

443 column. Samples were applied to a HiTrap 5 ml TALON Crude column [GE Healthcare] pre-
444 equilibrated in Buffer A [50 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 0.5 mM TCEP, 10 mM
445 imidazole]. The column was washed with 5 column volumes of Buffer A, with retained protein
446 eluted by the addition of 5 CV of Buffer B (50 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 0.5
447 mM TCEP, 300 mM imidazole). Affinity/solubility tags were removed by incubation with either
448 HRV-3C (RecQ1, WRN) or TEV protease (RecQ5).
449
450 *UvrD*
451 Purified recombinant *E. coli* UvrD was kindly provided by Dr. Mohan Rajasekaran (Sussex
452 Drug Discovery Centre, University of Sussex, UK).
453
454 **REAGENTS**
455 **Solutions**
456 Mg-ATP = 50 mM $MgCl_2$, 50 mM ATP
457 Mg-ADP = 50 mM $MgCl_2$, 50 mM ADP
458
459 **Oligonucleotides**
460 Reverse-phase purified oligonucleotides were purchased from either Kaneka Eurogentec S.A.
461 (Seraing, Belgium) or Eurofins Genomics Germany GmbH (Ebersberg, Germany).
462
463 *ssDNA-15mer:* 5'-CGTACCCGATGTGTT-3'
464
465 *ssDNA-20mer:* 5'-CGTACCCGATGTGTTCGTTC-3'
466
467 *Forked-50mer:*
468 A: 5'-XGAACGAACACATCGGGTACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'
469 B: 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTACCCGATGTGTTCGTTCY-3'
470
471 Where X and Y are the following modifications:
472 Unwinding assay; X = BHQ2 (Black Hole Quencher 2), Y = TAMRA (tetramethylrhodamine)
473 Gel-based assay; X = none; Y = TAMRA
474 Dye-displacement assay; X = none; Y = none
475
476 FORK-A and FORK-B were annealed at a concentration of 200 μM using a slow-cooling cycle
477 programmed into a PCR thermal cycler, in a buffer containing 20 mM HEPES.NaOH pH 7.5,
478 50 mM NaCl and 1 mM $MgCl_2$.
479

480 **Commercially sourced ML216**

481 ML216 was purchased from Merck KGaA (Darmstadt, Germany), product code: SML0661.
482 ML216-A was purchased from Cayman Chemical (Ann Arbor, Michigan USA), product code:
483 15186.

484

485 **BIOCHEMICAL ASSAYS**

486 **Fluorescence-based DNA unwinding assay**

487 Methodology is based on that previously reported by Rosenthal *et al*[14]. Briefly, assays were
488 carried out in 384-well black plates, with measurements taken at emission and excitation
489 wavelengths of 540 and 590 nm respectively, in a PHERAstar multimode plate reader [BMG
490 Labtech]. Assay buffer: 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM $MgCl_2$, 0.01% v/v Tween-
491 20, 2.5 µg/ml poly(dI-dC), 1 mM DTT.

492

493 28 µl of BLM-HD (at 0.535 nM in assay buffer) was pre-incubated with 2 µl of compound (2
494 mM stock dissolved in 100% v/v DMSO, over a range of final concentrations up to 100 µM)
495 for a period for 15 minutes at room temperature. Next, 10 µl of substrate (40 nM forked-50mer
496 dsDNA and 2000 µM Mg-ATP) was added, then incubated for a further 20 minutes at room
497 temperature, before the final fluorescent intensity for each well was measured.

498

499 Assay conditions (compounds **2** to **7** and ML216): 0.375 nM BLM-HD, 10 nM annealed DNA
500 substrate, 500 µM Mg-ATP in a reaction volume of 40 µl over a 20-minute incubation period.

501

502 Data for compound **1** are taken from an earlier iteration of the assay and were measured used
503 the conditions: 3.75 nM BLM-HD, 75 nM annealed DNA substrate, 120 µM Mg-ATP in a
504 reaction volume of 40 µl over a 20-minute incubation period.

505

506 **Malachite-Green ATP turnover assay**

507 Assay uses the PiColorLock Gold Phosphate Detection System from Innova Bioscience
508 following the manufacturer's recommended protocol. Briefly, assays were carried out in 96-
509 well clear flat-bottomed plates, with absorbance measurements taken at a wavelength of 630
510 nm in a CLARIOstar multimode plate reader [BMG Labtech]. Assay buffer: 50 mM Tris-HCl
511 pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, 0.05% v/v Tween-20, 0.5 mM TCEP.
512 165 µl of BLM-HD and ssDNA-20mer (at a concentration of 2.4 nM and 121 nM, respectively)
513 was pre-incubated with 10 µl of compound (2 mM stock dissolved in 100 % v/v DMSO, over a
514 range of final concentrations up to 100 µM) for a period of 15 minutes at room temperature.
515 Next, 25 µl of Mg-ATP substrate (at 16 mM) was added. After 20 minutes, reactions were

516 stopped by the addition of 50 μl Gold mix (a 100:1 ratio of PiColorLock:Accelerator reagents).
517 After 2 minutes, 20 μl of stabiliser solution was added to each well. After a further 30 minutes
518 absorbance measurements were taken.
519
520 Assay conditions: 2 nM BLM-HD, 100 nM ssDNA-20mer and 2 mM Mg-ATP in a reaction
521 volume of 200 µL over a 20-minute incubation period.
522
523 **Gel-based assay**
524 Assay buffer: 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM $MgCl_2$, 0.01% v/v Tween-20, 2.5
525 µg/ml poly(dI-dC), 1 mM DTT. 28 μl of BLM-HD (at a concentration of 2.9 nM) was pre-
526 incubated with 2 μl of compound (2 mM stock dissolved in 100 % v/v DMSO, over a range of
527 concentrations up to 100 µM) for a period of 15 minutes at room temperature. Next, 10 µl of
528 substrate (300 nM forked-50mer dsDNA and 4.8 mM Mg-ATP) were added. After 10 minutes,
529 reactions were terminated by the addition of 1 x loading dye (6 x solution: 10 mM Tris-HCl pH
530 7.5, 0.03% w/v bromophenol blue, 60% v/v glycerol, 60 mM EDTA). The samples were then
531 loaded onto a 15% native gel (29:1 acrylamide:bis-acrylamide, 0.5 x TBE), separated by
532 electrophoresis, and then visualized using a FLA-1500 fluorimager [Fujifilm, Bedford, UK]. The
533 intensity of each species on the gel was quantified using the analysis tools provided in the
534 software package Fiji[32].
535
536 **Topoisomerase I DNA-unwinding assay**
537 Assay uses the DNA Unwinding Assay Kit from Inspiralis (Norwich, UK) following the
538 manufacturer's recommended protocol. Resultant samples were applied to a 1% w/v agarose
539 gel (in 1 x TAE buffer), separated by electrophoresis, stained with ethidium bromide, and then
540 visualised with a UV-transilluminator / digital gel documentation system.
541
542 **Dye-displacement assay**
543 Fluorescence intensity was measured in a CLARIOstar multi-mode plate reader [BMG
544 Labtech] with excitation and emission wavelengths of 485 nm and 520 nm respectively, in
545 384-well black plates. 28 µl of forked-50mer dsDNA (at a final concentration of 800 nM) was
546 pre-incubated with 10 µl of SYBR Green II (1:200 dilution) for a period of 20 minutes at room
547 temperature. 2 µl of compound (1 mM stock dissolved in 100 % v/v DMSO, over a range of
548 final concentrations up to 50 µM) was then added. Measurements were taken after incubation
549 times of 20, 45 and 60 minutes. Assay buffer: 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM
550 $MgCl_2$, 0.01% v/v Tween-20, 1 mM DTT.
551

552 Assay conditions: 800 nM annealed DNA substrate and 1:800 SYBR Green II in a reaction
553 volume of 40 µL over a 20-minute incubation period.
554
555 **BIOPHYSICAL ASSAYS**
556 **Microscale thermophoresis (MST)**
557 Experiments were performed in a Monolith NT.115 instrument from NanoTemper
558 Technologies GmbH (München, Germany). Purified recombinant protein was labelled using
559 a Monolith NT RED-Maleimide Protein Labelling Kit supplied by the manufacturer, following
560 the recommended protocol. 19 µl of BLM-HD (at a final concentration of 75 nM) was mixed
561 with 1 µl of the required 'ligand' solution (ssDNA and / or compound) and incubated for 15
562 minutes at room temperature, before being transferred to 'premium' capillaries for
563 measurement. Experiments were performed at a temperature of 25 °C, with settings of 20%
564 excitation power, 20% MST power. Assay buffer: 50 mM Tris-HCl pH 7.5, 100 mM NaCl,
565 2 mM $MgCl_2$, 0.05% v/v Tween-20, 0.5 mM TCEP.
566
567 **CRYSTALLOGRAPHY**
568 *BLM-HD$^{\Delta WHD}$ / ADP*
569 Prior to setting up of crystallisation screens BLM-HD$^{\Delta WHD}$ at a concentration of 15 mg/ml was
570 combined with glycerol (100% v/v) and Mg-ADP (50 mM) to produce final concentrations of
571 10% v/v and 2 mM respectively. 150 nl of the prepared complex was combined with 150 nl of
572 crystallisation reagent in MRC2 sitting drop vapour diffusion experiments against a reservoir
573 volume of 50 µl. Crystals were obtained in condition A8 of the Morpheus HT-96 screen (0.06
574 M divalents, 37.5% Buffer System 2 and 37.5% Precipitant Mix 4; Molecular Dimensions
575 [Sheffield, UK] at 4°C after a period of approximately 1 week.
576
577 Divalents = 0.3M magnesium chloride, 0.3M calcium chloride
578 Buffer system 2 = 1M sodium HEPES, MOPS (acid) pH 7.5
579 75% Precipitant Mix 4 = 25% w/v MPD, 25% v/v PEG1000, 25% w/v PEG 3350
580 Cryoprotection for data collection was achieved by stepwise soaking of crystals in buffers
581 containing increasing amounts of ethylene glycol, to a final concentration of 20% (v/v).
582 Diffraction data to a resolution of 1.53 Angstrom were collected from a single crystal, on
583 beamline I04 at the Diamond Light Source (Didcot, UK). Crystals were in space group $P2_1$
584 with one molecule of BLM-HD$^{\Delta WHD}$ plus associated ligands forming the asymmetric unit.
585
586 *BLM-HD$^{\Delta WHD}$ / ADP / ssDNA-15mer / compound 2*

587 BLM-HD$^{\Delta WHD}$ was mixed with ssDNA-15mer at a 1:1.2 molar ratio (protein:DNA) to produce a
588 final concentration of 15 mg/ml with respect to protein. Compound **2** was then added to a final
589 concentration of 3 mM (from a stock at 100 mM in 100% v/v DMSO) and incubated with the
590 protein:DNA complex overnight at 4°C. Prior to setting up crystallisation trials the complex
591 was combined with glycerol (100% v/v) and Mg-ADP (50mM) to produce final concentrations
592 of 10% v/v and 2 mM respectively. 150 nl of the prepared complex was combined with 150 nl
593 of crystallisation reagent in MRC2 sitting drop vapour diffusion experiments against a reservoir
594 volume of 50 µl. Crystals were obtained in condition C9 of the Morpheus HT-96 screen (0.09
595 M NPS, 0.1M Buffer System, 30% Precipitant Mix 1, Molecular Dimensions) at 4°C after a
596 period of approximately 1 week.
597
598 NPS = 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulphate
599 Buffer System 1 = 1.0 M imidazole, MES monohydrate (acid) pH 6.5
600 60% Precipitant Mix 1 = 40% v/v PEG 500 MME, 20% w/v PEG 20000
601
602 Cryoprotection for data collection was achieved by stepwise soaking of crystals in buffers
603 containing increasing amounts of ethylene glycol, to a final concentration of 20% (v/v).
604 Diffraction data to a resolution of 3.0 Angstrom were collected from a single crystal, on
605 beamline I03 at the Diamond Light Source (Didcot, UK). Crystals were in space group P1 with
606 six molecules of BLM-HD$^{\Delta WHD}$ plus associated ligands forming the asymmetric unit.
607
608 **Data processing and model building**
609 Diffraction data were automatically processed at the synchrotron by the xia2 pipeline[33], using
610 software packages DIALS[34,35] or XDS [36] and Aimless[37]. For BLM-HD$^{\Delta WHD}$/ADP, coordinates
611 corresponding to the helicase domain were extracted from PDB entry 4O3M and provided as
612 a search model for molecular replacement using Phaser[38]. For BLM-HD$^{\Delta WHD}$/ADP/ssDNA-
613 15mer/compound 2, the rebuilt and refined coordinates for BLM-HD$^{\Delta WHD}$ were used as the
614 search model. Initial models were extended and improved by iterative rounds of building in
615 Coot[39] and refinement in either PHENIX[40] or BUSTER[41] to produce the final deposited models.
616 Crystallisation and refinement statistics are provided in Supplementary Table 1.

617 **Data plotting and analysis**

618 All experimental data were plotted and analysed using GraphPad Prism[42].

619

620 **REFERENCES**

621 1. Croteau, D.L., Popuri, V., Opresko, P.L. & Bohr, V.A. Human RecQ helicases in DNA
622 repair, recombination, and replication. *Annu Rev Biochem* **83**, 519-52 (2014).
623 2. Wu, Y. Unwinding and rewinding: double faces of helicase? *J Nucleic Acids* **2012**,
624 140601 (2012).
625 3. Urban, V., Dobrovolna, J. & Janscak, P. Distinct functions of human RecQ helicases
626 during DNA replication. *Biophys Chem* **225**, 20-26 (2017).
627 4. Chan, E.M. et al. WRN helicase is a synthetic lethal target in microsatellite unstable
628 cancers. *Nature* **568**, 551-556 (2019).
629 5. Lieb, S. et al. Werner syndrome helicase is a selective vulnerability of microsatellite
630 instability-high tumor cells. *Elife* **8**(2019).
631 6. Chan, K.L., North, P.S. & Hickson, I.D. BLM is required for faithful chromosome
632 segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J*
633 **26**, 3397-409 (2007).
634 7. Bernstein, K.A., Gangloff, S. & Rothstein, R. The RecQ DNA helicases in DNA repair.
635 *Annu Rev Genet* **44**, 393-417 (2010).
636 8. Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B. & Pearl, F.M. Therapeutic
637 opportunities within the DNA damage response. *Nat Rev Cancer* **15**, 166-80 (2015).
638 9. Aggarwal, M. & Brosh, R.M., Jr. Hitting the bull's eye: novel directed cancer therapy
639 through helicase-targeted synthetic lethality. *J Cell Biochem* **106**, 758-63 (2009).
640 10. Wang, H. et al. BLM prevents instability of structure-forming DNA sequences at
641 common fragile sites. *PLoS Genet* **14**, e1007816 (2018).
642 11. Behan, F.M. et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9
643 screens. *Nature* **568**, 511-516 (2019).
644 12. Kategaya, L., Perumal, S.K., Hager, J.H. & Belmont, L.D. Werner Syndrome Helicase Is
645 Required for the Survival of Cancer Cells with Microsatellite Instability. *iScience* **13**,
646 488-497 (2019).
647 13. Nguyen, G.H. et al. A small molecule inhibitor of the BLM helicase modulates
648 chromosome stability in human cells. *Chem Biol* **20**, 55-62 (2013).
649 14. Rosenthal, A.S. et al. Discovery of ML216, a Small Molecule Inhibitor of Bloom (BLM)
650 Helicase. in *Probe Reports from the NIH Molecular Libraries Program* (Bethesda
651 (MD), 2010).
652 15. Yin, Q.K. et al. Discovery of Isaindigotone Derivatives as Novel Bloom's Syndrome
653 Protein (BLM) Helicase Inhibitors That Disrupt the BLM/DNA Interactions and
654 Regulate the Homologous Recombination Repair. *J Med Chem* **62**, 3147-3162 (2019).
655 16. Del Villar-Guerra, R., Gray, R.D., Trent, J.O. & Chaires, J.B. A rapid fluorescent
656 indicator displacement assay and principal component/cluster data analysis for
657 determination of ligand-nucleic acid structural selectivity. *Nucleic Acids Res* **46**, e41
658 (2018).
659 17. Tse, W.C. & Boger, D.L. A fluorescent intercalator displacement assay for establishing
660 DNA binding selectivity and affinity. *Acc Chem Res* **37**, 61-9 (2004).

661 18. Newman, J.A. et al. Crystal structure of the Bloom's syndrome helicase indicates a
662 role for the HRDC domain in conformational changes. *Nucleic Acids Res* **43**, 5221-35
663 (2015).
664 19. Swan, M.K. et al. Structure of human Bloom's syndrome helicase in complex with
665 ADP and duplex DNA. *Acta Crystallogr D Biol Crystallogr* **70**, 1465-75 (2014).
666 20. Manthei, K.A., Hill, M.C., Burke, J.E., Butcher, S.E. & Keck, J.L. Structural mechanisms
667 of DNA binding and unwinding in bacterial RecQ helicases. *Proc Natl Acad Sci U S A*
668 **112**, 4292-7 (2015).
669 21. Zittel, M.C. & Keck, J.L. Coupling DNA-binding and ATP hydrolysis in Escherichia coli
670 RecQ: role of a highly conserved aromatic-rich sequence. *Nucleic Acids Res* **33**, 6982-
671 91 (2005).
672 22. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast
673 protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*
674 **60**, 2256-68 (2004).
675 23. Voter, A.F., Qiu, Y., Tippana, R., Myong, S. & Keck, J.L. A guanine-flipping and
676 sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nat*
677 *Commun* **9**, 4201 (2018).
678 24. Newman, J.A., Aitkenhead, H., Savitsky, P. & Gileadi, O. Insights into the RecQ
679 helicase mechanism revealed by the structure of the helicase domain of human
680 RECQL5. *Nucleic Acids Res* **45**, 4231-4243 (2017).
681 25. Morozov, V., Mushegian, A.R., Koonin, E.V. & Bork, P. A putative nucleic acid-binding
682 domain in Bloom's and Werner's syndrome helicases. *Trends Biochem Sci* **22**, 417-8
683 (1997).
684 26. Kim, Y.M. & Choi, B.S. Structure and function of the regulatory HRDC domain from
685 human Bloom syndrome protein. *Nucleic Acids Res* **38**, 7764-77 (2010).
686 27. Gyimesi, M., Sarlos, K. & Kovacs, M. Processive translocation mechanism of the
687 human Bloom's syndrome helicase along single-stranded DNA. *Nucleic Acids Res* **38**,
688 4404-14 (2010).
689 28. Gorbalenya, A.E. & Koonin, E.V. Helicases - Amino-Acid-Sequence Comparisons and
690 Structure-Function-Relationships. *Current Opinion in Structural Biology* **3**, 419-429
691 (1993).
692 29. Hall, M.C. & Matson, S.W. Helicase motifs: the engine that powers DNA unwinding.
693 *Mol Microbiol* **34**, 867-77 (1999).
694 30. Kitano, K. Structural mechanisms of human RecQ helicases WRN and BLM. *Front*
695 *Genet* **5**, 366 (2014).
696 31. Wu, L. et al. The HRDC domain of BLM is required for the dissolution of double
697 Holliday junctions. *EMBO J* **24**, 2679-87 (2005).
698 32. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat*
699 *Methods* **9**, 676-82 (2012).
700 33. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta Crystallogr D*
701 *Biol Crystallogr* **69**, 1260-73 (2013).
702 34. Beilsten-Edmands, J. et al. Scaling diffraction data in the DIALS software package:
703 algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr D Struct*
704 *Biol* **76**, 385-399 (2020).
705 35. Winter, G. et al. DIALS: implementation and evaluation of a new integration package.
706 *Acta Crystallogr D Struct Biol* **74**, 85-97 (2018).
707 36. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).

37. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).
38. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
40. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
41. Bricogne, G. et al. BUSTER version 2.10.3. *Global Phasing Ltd., Cambridge, UK* (2020).
42. GraphPad Prism, v. *GraphPad Software Inc., La Jolla, CA, USA*.
43. Laskowski, R.A. & Swindells, M.B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-86 (2011).
44. Rosenthal, A.S. et al. Synthesis and SAR studies of 5-(pyridin-4-yl)-1,3,4-thiadiazol-2-amine derivatives as potent inhibitors of Bloom helicase. *Bioorg Med Chem Lett* **23**, 5660-6 (2013).
45. Schrödinger, L. The PyMOL Molecular Graphics System, Version 2.0.
46. PDBsum. https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html.

ACKNOWLEDGMENTS

General: We thank Will Pearce (Sussex Drug Discovery Centre) for discussions relating to assay development. We are also grateful to the Diamond Light Source Ltd., Didcot, UK for access to synchrotron radiation.

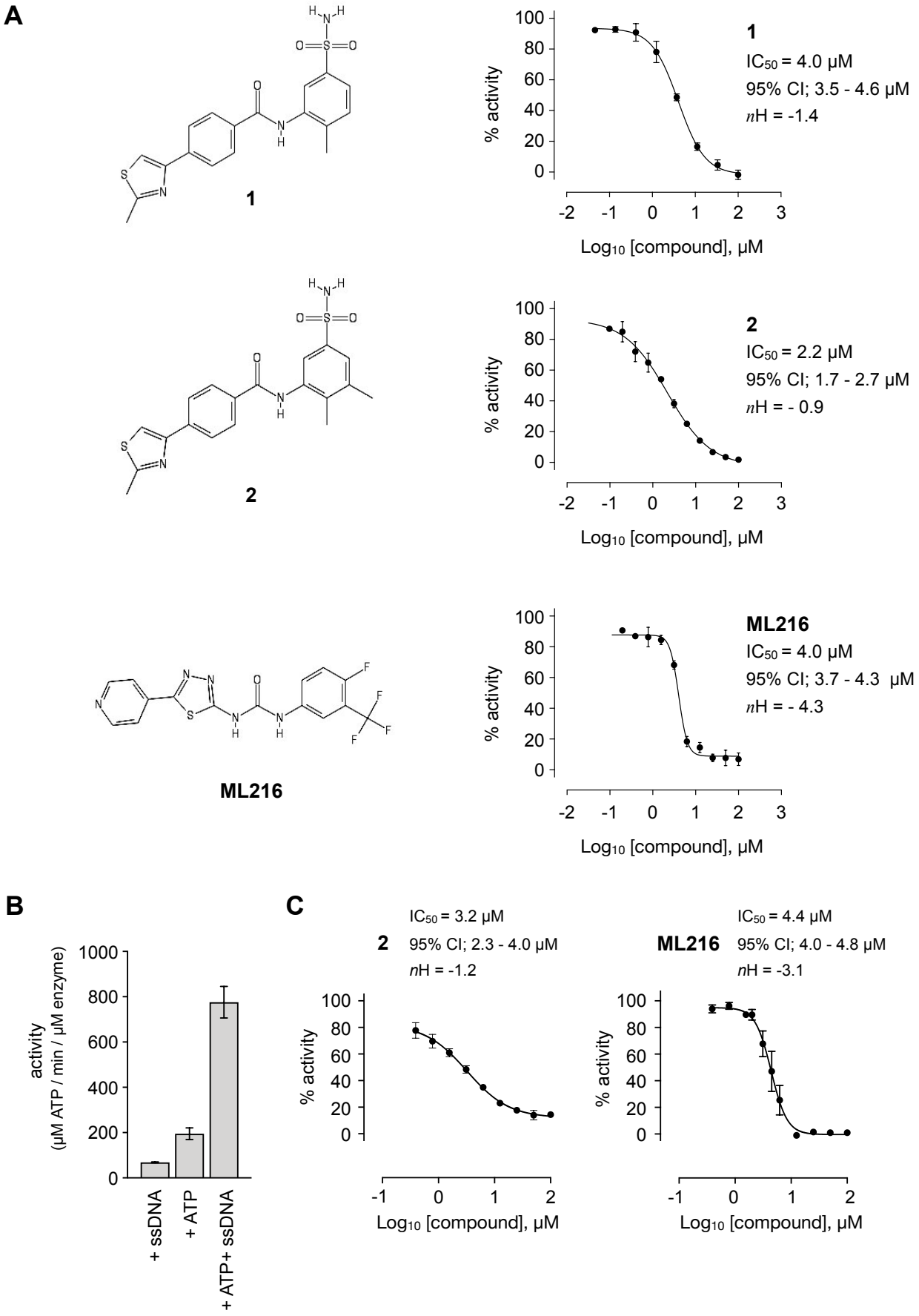
Funding: This work was supported by funding from: Chinese Scholarship Council (XC and FGMP), Wellcome Trust 110578/Z/15/Z (SEW), Cancer Research UK Programme Grant C302/A24386 (LHP and AWO).

Competing Interests: The authors declare no competing interests.

Author Contributions: Conceptualisation: SEW, LHP, FGMP, AWO; Methodology: GW, SW, JJRH, AWO; Investigation: XC, YA, CELF, RA-B, JRB, SMR, SEW, LHP, FGMP, AWO; Writing — Original Draft: AWO; Writing — Review and Editing: XC, LHP, AWO; Visualisation: AWO; Supervision: GW, SW, SEW, FGMP, AWO; Funding Acquisition: XC, FGMP, SEW, LHP, AWO.

Data and materials availability: Coordinates and structures factors have been deposited in the Protein Data Bank (PDB) with accession codes 7AUC and 7AUD.

Figure 1


A
1
IC50 = 4.0 μM
95% CI; 3.5 - 4.6 μM
nH = -1.4
% activity
Log10 [compound], μM
2
IC50 = 2.2 μM
95% CI; 1.7 - 2.7 μM
nH = - 0.9
ML216
IC50 = 4.0 μM
95% CI; 3.7 - 4.3 μM
nH = - 4.3
B
activity
(μM ATP / min / μM enzyme)
+ ssDNA
+ ATP
+ ATP+ ssDNA
C
2
IC50 = 3.2 μM
95% CI; 2.3 - 4.0 μM
nH = -1.2
ML216
IC50 = 4.4 μM
95% CI; 4.0 - 4.8 μM
nH = -3.1

748 **Figure 1. Inhibition of BLM helicase unwinding activity**

749 (A, left) Chemical drawings for compounds **1**, **2** and ML216. (A, right) Dose response curves
750 from fluorescence-based DNA unwinding assays with BLM-HD. Experimental data were fitted
751 with a four parameter, log(inhibitor) vs. response model with variable slope. Calculated values
752 for IC_{50}, Hill slope (*n*H) and 95% confidence intervals (95% CI) are given in each case. (B)
753 Turnover of ATP by BLM-HD, as measured by a malachite-green end-point assay, is strongly
754 stimulated in the presence of a 20-base single-stranded oligonucleotide. (C) Dose response
755 curves from ATP-turnover assays with BLM-HD. Data were fitted as for (A). In each case data
756 points are the mean of 3 technical replicates, with error bars representing one standard
757 deviation (1 SD).

Figure 2

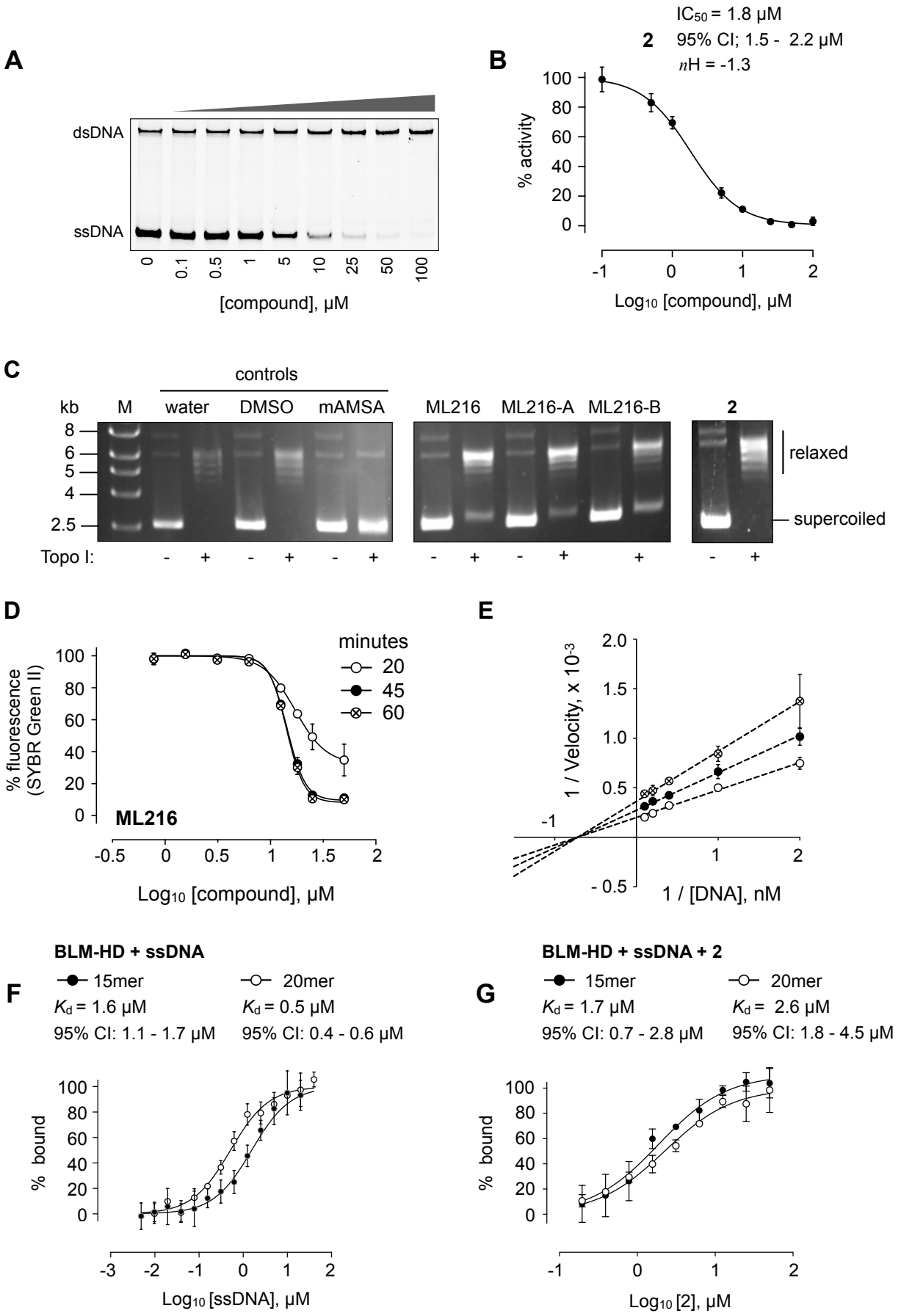
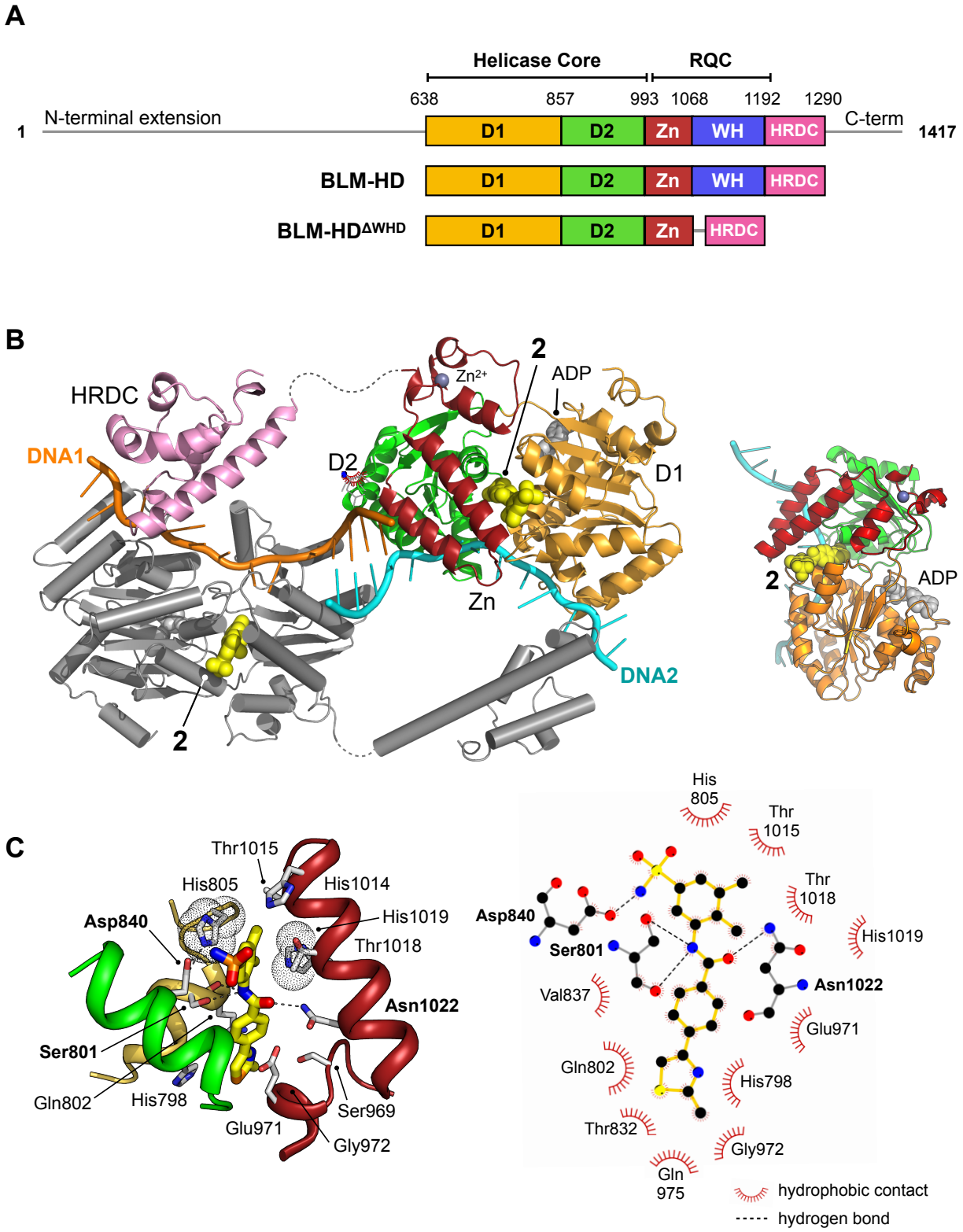

A
dsDNA
ssDNA
0
0.1
0.5
1
5
10
25
50
100
[compound], μM
B
IC50 = 1.8 μM
2
95% CI; 1.5 - 2.2 μM
nH = -1.3
% activity
Log10 [compound], μM
C
controls
kb
M
water
DMSO
mAMSA
ML216
ML216-A
ML216-B
2
8
6
5
4
2.5
relaxed
supercoiled
Topo I:
D
minutes
20
45
60
% fluorescence (SYBR Green II)
ML216
Log10 [compound], μM
E
1 / Velocity, x 10-3
1 / [DNA], nM
F
BLM-HD + ssDNA
15mer
20mer
Kd = 1.6 μM
Kd = 0.5 μM
95% CI: 1.1 - 1.7 μM
95% CI: 0.4 - 0.6 μM
% bound
Log10 [ssDNA], μM
G
BLM-HD + ssDNA + 2
15mer
20mer
Kd = 1.7 μM
Kd = 2.6 μM
95% CI: 0.7 - 2.8 μM
95% CI: 1.8 - 4.5 μM
% bound
Log10 [2], μM

Figure 2. DNA interaction assays

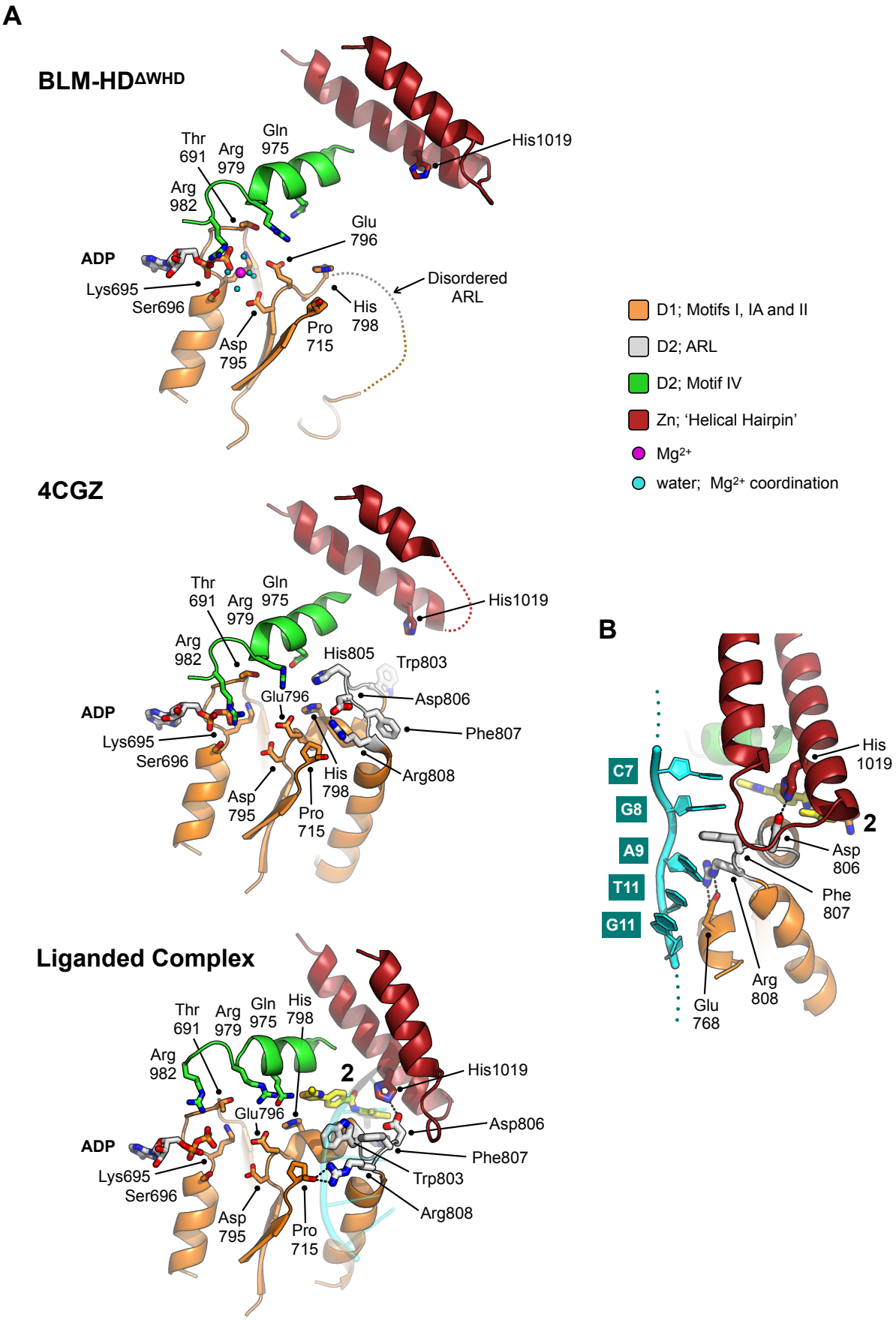
758 (A) Titration of BLM-HD with **2** prevents the unwinding of a forked-50mer dsDNA substrate
759 into its component strands, as judged by native gel electrophoresis**.** (B) Quantification of
760 inhibitory activity by **2** in the gel-based activity assay. Experimental data were fitted with a
761 four parameter, log(inhibitor) vs. response model with variable slope. Calculated values for
762 IC_{50}, *n*H and 95% CI are given in each case. (C) Representative results from a Topoisomerase
763 I (Topo I) DNA-unwinding assay. M = molecular mass maker; DMSO = buffer supplemented
764 with dimethyl sulfoxide control; mAMSA = mAmsacrine; ML216, ML216-A, ML216B = refer to
765 the three independent sources of the compound as described in the main text of the
766 manuscript **(D)** Dose response curves from SYBR-Green II dye displacement assays, using a
767 forked-50mer DNA duplex incubated with ML216 for a period of 20 (open circles), 45 (filled
768 circles) and 60 minutes (crossed circles). Fitted lines are intended as visual aids only. **(E)**
769 Lineweaver-Burk plot for data generated at three compound concentrations (0, 5 and 10 μM)
770 in a colourimetric ATP turnover assay. Linear regression produces an intercept of all data on
771 the X-axis indicating that **2** is a non-competitive inhibitor (i.e. same K_m, altered V_{max}
772 parameter). (**F, G**) Binding isotherms for binding of BLM-HD to ssDNA-15mer and -20mer, or
773 to compound **2** in the presence of either oligonucleotide, as determined by microscale
774 thermophoresis (MST). Experimental data were fitted with a one-site, specific binding model.
775 Values for K_d and 95% CI are given in each case. For all plots, data represent the mean of
776 three technical replicates with error bars representing 1 SD.

Figure 3


A
Helicase Core
RQC
638
857
993
1068
1192
1290
1
N-terminal extension
D1
D2
Zn
WH
HRDC
C-term
1417
BLM-HD
BLM-HDΔWHD
B
HRDC
DNA1
D2
Zn2+
2
ADP
D1
Zn
DNA2
2
2
ADP
C
Thr1015
His805
His1014
Asp840
His1019
Thr1018
Asn1022
Ser801
Gln802
His798
Glu971
Ser969
Gly972
His 805
Thr 1015
Thr 1018
Asp840
His1019
Ser801
Asn1022
Val837
Glu971
Gln802
His798
Thr832
Gly972
Gln 975
hydrophobic contact
hydrogen bond

777 **Figure 3. Structural overview of BLM-HD$^{\Delta WHD}$ in complex with compound 2**
778 (A) Schematic representation of the domain composition and respective amino acid
779 boundaries for full-length human BLM and the two expression constructs used in this study
780 BLM-HD and BLM-HD$^{\Delta WHD}$. D1 and D2 = domains 1 and 2 of the helicase core; Zn = zinc-
781 binding domain; WH = winged-helix; HRDC = Helicase and RNaseD C-terminal domain; RQC
782 = RecQ C-terminal domain. (B and B inset) Molecular secondary structure cartoon highlighting
783 components of a 'pseudo-symmetrical' dimer found in the asymmetric unit of BLM-HD$^{\Delta WHD}$ /
784 Mg-ADP / **2** / ssDNA crystals**,** driven by partial complementarity of the single-stranded 15mer
785 oligonucleotide at its 5' end (DNA1 and DNA2, coloured orange and cyan respectively).
786 Compound **2** (yellow-coloured spheres) binds to a small pocket found on the opposite side to
787 that which binds nucleotide (grey-coloured spheres). (C, left) Molecular cartoon representation
788 highlighting interactions made between **2** and BLM-HD$^{\Delta WHD}$. Key amino acid residues are
789 labelled and shown in stick representation, with carbon atoms coloured according to the
790 schematic shown in panel A. Compound **2** is shown in stick representation, with carbon atoms
791 coloured yellow. Potential hydrogen bonds are indicated by black dotted lines. (C, right)
792 Modified LIGPLOT+[43] diagram of protein-compound interactions. See associated key for
793 additional detail.

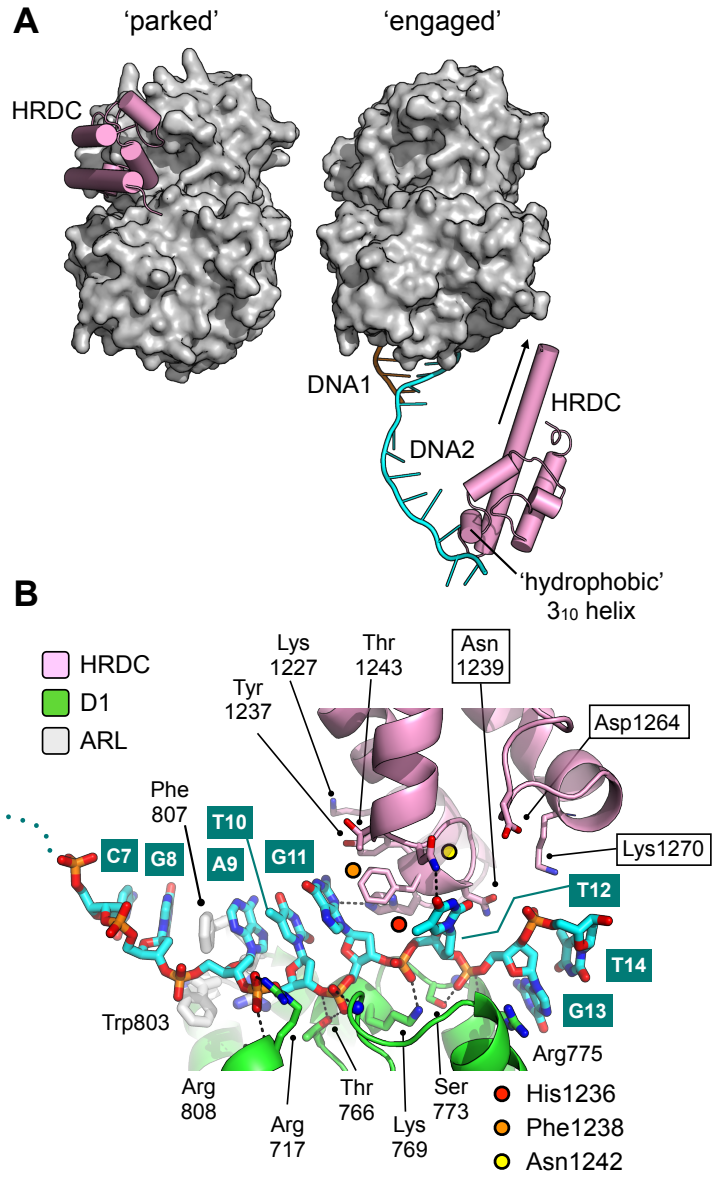
Figure 4


A
BLM-HDΔWHD
Thr 691
Arg 979
Gln 975
Arg 982
His1019
Glu 796
ADP
Lys695
Ser696
Disordered ARL
His 798
Pro 715
Asp 795
D1; Motifs I, IA and II
D2; ARL
D2; Motif IV
Zn; 'Helical Hairpin'
Mg2+
water; Mg2+ coordination
4CGZ
Thr 691
Arg 979
Gln 975
Arg 982
His1019
His805
Trp803
Glu796
Asp806
ADP
Phe807
Lys695
Ser696
His 798
Arg808
Asp 795
Pro 715
B
C7
G8
A9
T11
G11
His 1019
2
Asp 806
Phe 807
Arg 808
Glu 768
Liganded Complex
Thr 691
Arg 979
Gln 975
His 798
Arg 982
2
His1019
Glu796
Asp806
ADP
Phe807
Lys695
Ser696
Trp803
Arg808
Asp 795
Pro 715

794 **Figure 4. Structural transitions around the aromatic rich loop**

795 (A) Molecular secondary structure cartoons for the region surrounding the aromatic rich loop
796 (ARL) of BLM-HD$^{\Delta WHD}$ (top), PDB entry 4CGZ; BLM-HD in complex with DNA (middle) and
797 liganded complex; BLM-HD$^{\Delta WHD}$ in complex with ADP, ssDNA-15mer and **2** (bottom). The side
798 chains for key amino acid residues are shown in stick representation, with carbon atoms
799 coloured according to their respective domains (see associated key). Bound ADP and **2** are
800 also shown in stick representation, with carbon atoms coloured grey and yellow respectively.
801 (B) Expanded and rotated view highlighting the interactions made between the ARL and
802 ssDNA-15mer oligonucleotide (cartoon coloured cyan) in the liganded complex, also showing
803 the relative position of compound **2**. Potential hydrogens bonds are represented by black
804 dotted lines.

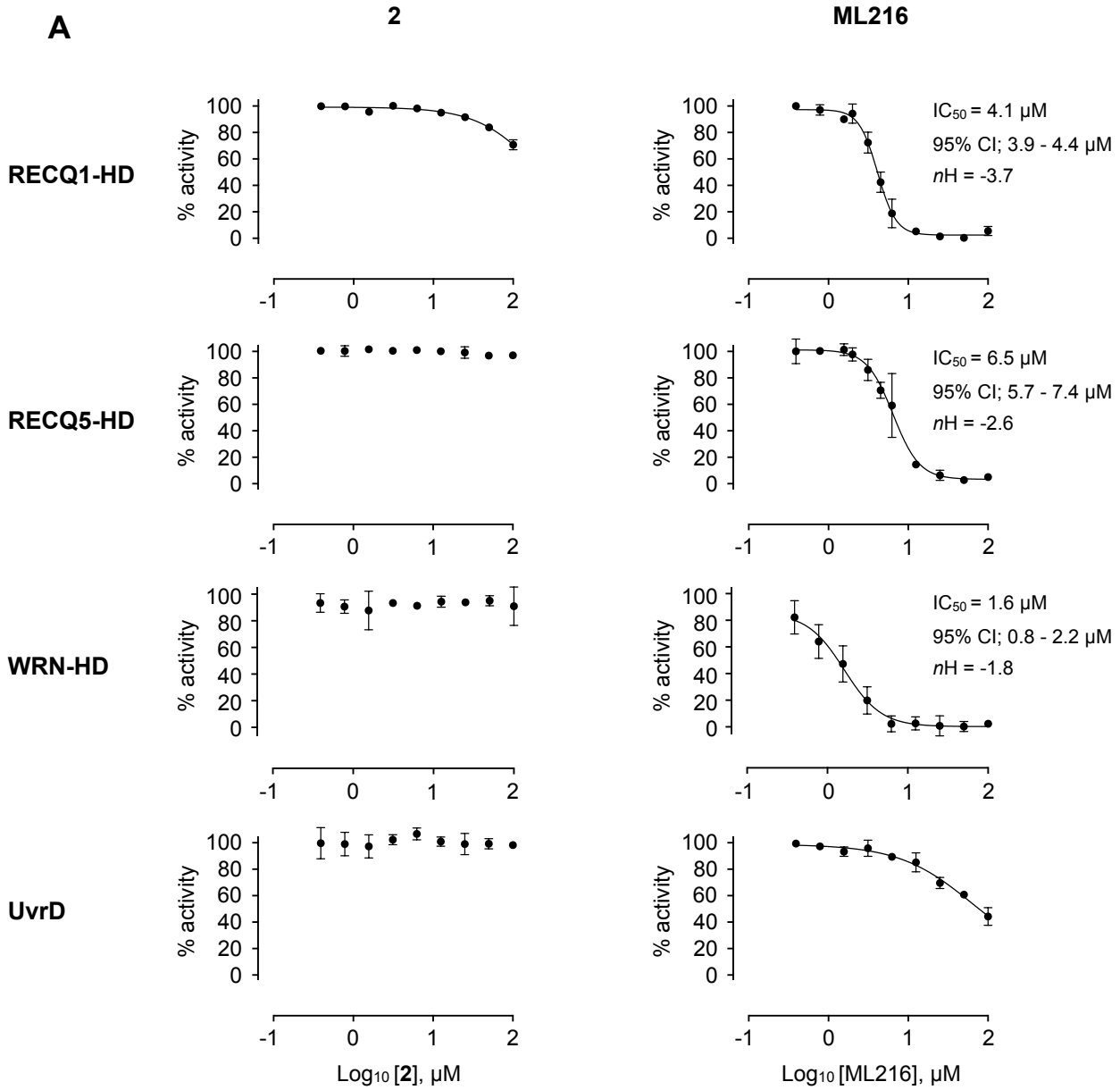
Figure 5

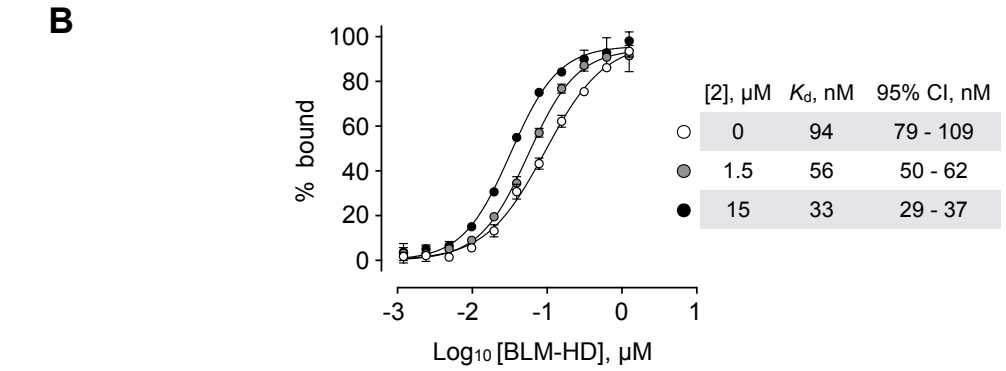

A
'parked'
'engaged'
HRDC
DNA1
HRDC
DNA2
'hydrophobic'
3_{10} helix
B
HRDC
D1
ARL
Lys
1227
Thr
1243
Asn
1239
Tyr
1237
Asp1264
Phe
807
T10
C7
G8
A9
G11
Lys1270
T12
T14
G13
Trp803
Arg775
Arg
808
Arg
717
Thr
766
Lys
769
Ser
773
His1236
Phe1238
Asn1242

805 **Figure 5. Repositioning of the HRDC domain**

806 (A) Molecular surface representation of BLM-HD$^{\Delta WHD}$ (left) and the liganded complex (right)
807 highlighting the relative positions of the HRDC domain (cylindrical helices coloured in pink).
808 The HRDC moves from a 'parked' position located on one side of the helicase core, to an
809 'engaged' position on the other side in order to interact with the bound ssDNA-15mer. The N-
810 terminus of the first HRDC alpha-helix is extended at the by ~ 6 aa, relative to the 'parked'
811 position (as indicated by an arrow). (B) Molecular secondary structure cartoon highlighting
812 interactions made by the HRDC to the bound ssDNA-15mer oligonucleotide. Side chains for
813 key amino acid residues are shown in stick representation, with carbon atoms coloured
814 according to their respective domains (see associated key). The bound ssDNA-15mer
815 oligonucleotide is involved in interactions with both the D1 domain (carbon atoms coloured
816 green) and the HRDC (carbon atoms coloured pink). Bound oligonucleotide is shown in stick
817 representation, with carbon atoms coloured cyan.

Figure 6


A
2
ML216
RECQ1-HD
RECQ5-HD
WRN-HD
UvrD
% activity
IC50 = 4.1 μM
95% CI; 3.9 - 4.4 μM
nH = -3.7
IC50 = 6.5 μM
95% CI; 5.7 - 7.4 μM
nH = -2.6
IC50 = 1.6 μM
95% CI; 0.8 - 2.2 μM
nH = -1.8
Log10 [2], μM
Log10 [ML216], μM


B
% bound
Log10 [BLM-HD], μM
[2], μM Kd, nM 95% CI, nM
0 94 79 - 109
1.5 56 50 - 62
15 33 29 - 37

818 **Figure 6. Selectivity profile of compound 2**

819 (A) Dose response curves from ATP-turnover assays for titration of compounds **2** and ML216
820 against purified recombinant BLM-HD, WRN-HD, RecQ1-HD, RecQ5-HD and UvrD
821 respectively. Calculated values for IC_{50}, *n*H and 95% CI are given in each case. **(B)** MST-
822 derived binding isotherms for the interaction of BLM-HD with ssDNA-20mer in the presence
823 of increasing concentrations of **2**. Calculated values for K_d and 95% CI are given in each case.
824 For all plots, data represent the mean of three technical replicates with error bars representing
825 1 SD.

Figure 7

A

motif I / ARL

```
                 * 798   * 801 * 802   * 805
BLM     734-IDEAHCVSQWGHDFR-808
WRN     667-VDEAHCISEWGHDFR-681
RECQ1   218-VDEVHCCSQWGHDFR-232
RECQ5   156-VDEAHCVSQWGHDFR-170
cons         !DEaHC.S#WGHDFR
```

motif III

```
                 * 832   * 837   * 840
829-LTATANPRVQKD-840
702-LTATASSSIRED-713
253-LTATATNHVLTD-264
191-LTATATPQVQED-202
    LTATAt..!..D
```

pre-motif VI

```
                 * 971 * 972   * 975
BLM     966-LPKSVEGYYQ-975
WRN     843-APKDMESYYQ-850
RECQ1   391-MSKSMENYYQ-400
RECQ5   336-IAKSMAGYYQ-345
cons         ..KsmegYYQ
```

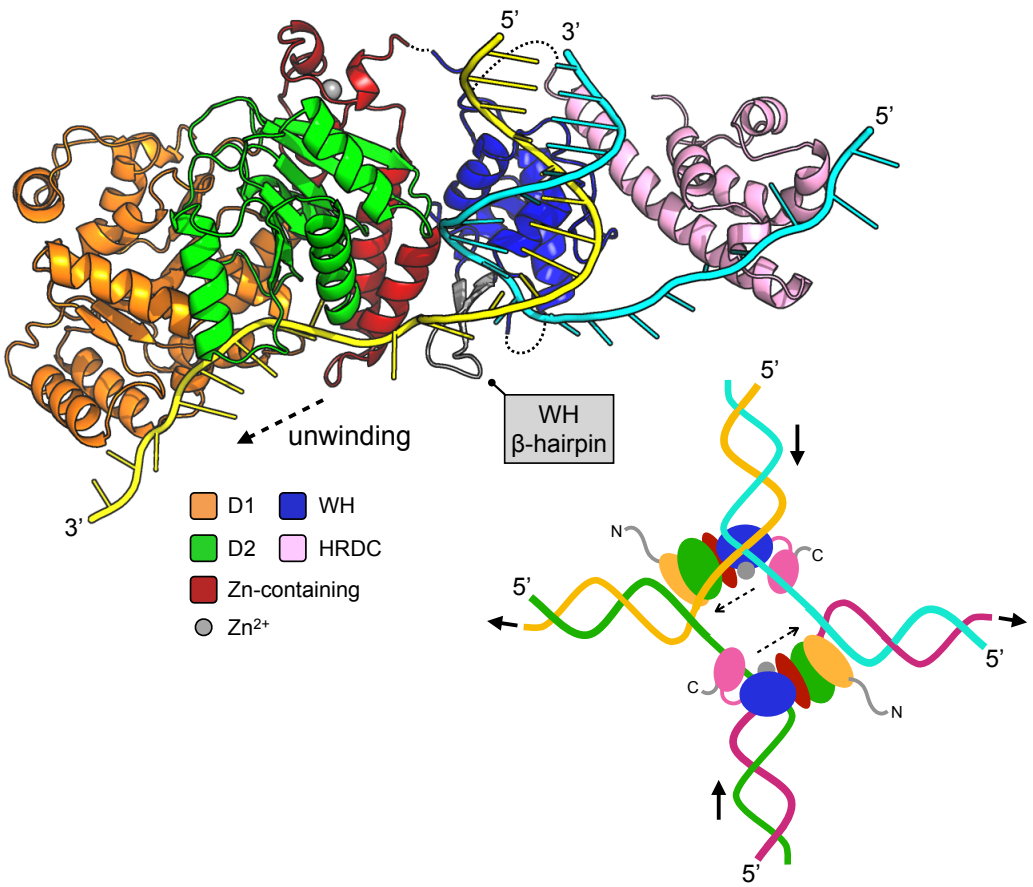
Zn / helical hairpin

```
                 * 1014 * 1015   * 1018 * 1019   * 1022
1004-LIMME-kdg---NHHTRETHFNNL-1023
 879-LLT-ir----NEKFRLYKLKMM-896
 429-MVV---men---------VGQQKL-440
 378-EVAKL-qekrgnka-SDKATIMAF-399
     lv...-------n...r.....l
```

α ····· α

hydrophobic contact
hydrogen bond
K high consensus
r low consensus
! isoleucine or leucine (I/L)
glutamic acid or glutamine (E/Q)

B



826 **Figure 7. Compound selectivity / Speculative model**

827 (A) The observed selectivity of compound **2** appears to arise from interactions made with
828 amino acids of the Zn-binding domain 'Helical Hairpin' that are poorly conserved (or absent)
829 from the other RecQ-family helicases. For each multiple amino acid sequence alignment
830 shown, highly conserved amino acids are coloured in red. Those conserved in at least two
831 RecQ family members (low consensus) are coloured blue. Amino acids residues of human
832 BLM involved in interactions with **2** are additionally highlighted. Please see associated key
833 for additional information. (B) A speculative model for how the HRDC domain may contribute
834 to the unwinding of a DNA duplex or a Holliday junction (inset) via transient interactions with
835 the passive or 'inactive' strand. See associated key for additional details.

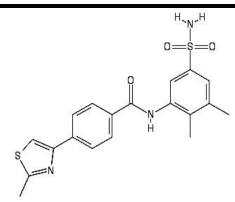
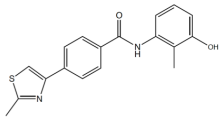
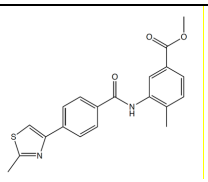
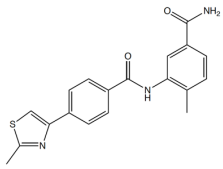
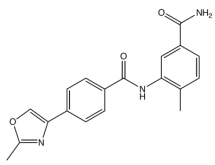
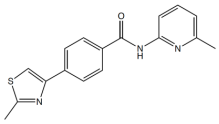
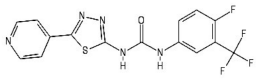
		unwinding	turnover
#	Chemical Drawing	IC_{50} [95% CI]; μM	IC_{50} [95% CI]; μM
2		2.2 [1.7 – 2.7]	3.2 [2.3 – 4.0]
3		3.5 [2.4 – 5.2]	5.3 [4.7 – 6.1]
4		6.6 [3.4 – 12.7]	11.2 [8.3 – 15.3]
5		12.8 [4.5 – 36.8]	47.86 [18.28 – 180.7]
6		56.9 [25.4 – 171.3]	40.94 [15.9 – 139.8]
7		No inhibition	No inhibition
ML216		4.0 [3.7 – 4.3]	4.4 [4.0 – 4.8]

Table 1. Summary of inhibition data for seven exemplars from the identified compound series IC_{50} values were determined by fitting of experimental data to log (inhibitor) vs response models provided in GraphPad Prism. Data for the unwinding assay correspond to three technical replicates from a single experiment. For the ATP-turnover assay data correspond to at least two independent experiments, each containing three technical replicates.

SUPPLEMENTARY MATERIALS

Synthetic Preparation of Test Molecules

All reactions were conducted under an atmosphere of nitrogen unless otherwise stated. Anhydrous solvents were used as purchased or were purified under nitrogen as follows using activated molecular sieves. Thin layer chromatography was performed on glass plates pre-coated with Merck silica gel 60 F254. Visualisation was achieved with U.V. fluorescence (254 nm) or by staining with a phosphomolybdic acid dip or a potassium permanganate dip. Flash column chromatography was carried out using pre-packed columns filled with Aldrich silica gel (40-63 μm) on an ISCO Combiflash Rf, or a Biotage Isolera Prime. Proton nuclear magnetic resonance spectra were recorded at 500 MHz on a Varian 500 spectrometer (at 30 °C), using residual isotopic solvent ($CHCl_3$, δH = 7.27 ppm, DMSO δH = 2.50 ppm, 3.33 ppm (H_2O)) as an internal reference. Chemical shifts are quoted in parts per million (ppm). Coupling constants (J) are recorded in Hertz (Hz). Carbon nuclear magnetic resonance spectra were recorded at 125 MHz on a Varian 500 spectrometer and are proton decoupled, using residual isotopic solvent ($CHCl_3$, δC = 77.00 ppm, DMSO δC = 39.52 ppm) as an internal reference. Carbon spectra assignments are supported by HSQC and DEPT editing and chemical shifts (δC) are quoted in ppm. Infrared spectra were recorded on a Perkin Elmer FT-IR One spectrometer as either an evaporated film or liquid film on sodium chloride plates. Absorption maxima are reported in wave numbers (cm^{-1}). Only significant absorptions are presented in the data, with key stretches identified in brackets. LCMS data was recorded on a Waters 2695 HPLC using a Waters 2487 UV detector and a Thermo LCQ ESI-MS. Samples were eluted through a Phenomenex Lunar 3μ C18 50 mm × 4.6 mm column, using acetonitrile and water acidified by 0.01% formic acid in three methods: method 1 (3:7 to 7:3 acetonitrile and water over 7 minutes), method 2 (3:7 to 7:3 acetonitrile and water over 4 minutes) and method 3 (19:1 to 1:19 acetonitrile and water over 10 minutes), High resolution mass spectrometry (HRMS) spectra were recorded on Bruker Daltonics Apex III ESI-MS, with an Apollo ESI probe using a methanol spray. Only molecular ions, fractions from molecular ions and other major peaks are reported as mass/charge (m/z) ratios.

***N*-(2-methyl-5-sulfamoyl-phenyl)-4-(2-methylthiazol-4-yl)benzamide (1)**

Methyl 4-(2-bromoacetyl)benzoate

A solution of methyl 4-acetylbenzoate (1.00 g, 5.61 mmol) and *p*-toluenesulfonic acid monohydrate (54 mg, 0.28 mmol) in acetonitrile (30 mL) was treated with *N*-bromosuccinimide (0.99 g, 5.61 mmol) and the reaction mixture heated to 80 °C for 16 h. The solvent was removed under reduced pressure. The crude product was taken up in saturated aq. $NaHCO_3$ (15 mL) and extracted with ethyl acetate (3 × 15 mL). The combined organic components were then washed with brine (15 mL), dried over $MgSO_4$, filtered and concentrated under reduced

873 pressure. The crude product was purified by column chromatography (silica 24 g, 0 to 10%
874 ethyl acetate in petroleum ether) to yield the desired compound as a light yellow solid (1.12 g,
875 70%). Rf 0.66 (petroleum ether:ethyl acetate, 9:1); ^{1}H-NMR (500 MHz, DMSO-d6) δ 8.14 –
876 8.01 (m, 4H, H-2, H-3), 4.98 (s, 2H, $COCH_2Br$), 3.89 (s, 3H, $COOCH_3$).
877 *Methyl 4-(2-methylthiazol-4-yl)benzoate*
878 To methyl 4-(2-bromoacetyl)benzoate (2.12 g, 8.23 mmol) in *N,N*-dimethylformamide (30 mL)
879 was added thioacetamide (931 mg, 12.4 mmol) and the reaction mixture stirred at ambient
880 temperature for 16 h. Upon completion, water was added to the reaction mixture. The resulting
881 precipitate was collected by vacuum filtration and dried under reduced pressure to afford the
882 desired compound as a white solid (1.48 g, 73%). m.p. 223–225 °C; ^{1}H-NMR (500 MHz,
883 DMSO-d6) δ 8.14 (s, 1H, H-6), 8.08 (d, J 8.2, 2H, H-2), 8.01 (d, J 8.2, 2H, H-3), 3.87 (s, OCH_3),
884 2.73 (s, CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.5 (CO), 166.4 (C-8), 153.0 (C-5), 138.9
885 (C-4), 130.2 (C-2), 129.1 (C-1), 126.5 (C-3), 116.8 (C-6), 52.5 (OCH_3), 19.4 (CH_3); IR (neat,
886 v_{max}, cm^{-1}) 3106, 2943, 1713, 1606, 1436, 1409, 1270, 1170; LCMS (LCQ) Rt = 2.9 min
887 (method 2), m/z (ESI+) 234.2 $[M+H]^+$; HRMS m/z (ESI): calcd. for $C_{12}H_{11}NO_2S$ $[M+H]^+$
888 234.0583, found 234.0583.
889 *4-(2-Methylthiazol-4-yl)benzoic acid*
890 To methyl 4-(2-methylthiazol-4-yl)benzoate (1.46 g, 6.26 mmol) in methanol (24 mL) and water
891 (8 mL) was added sodium hydroxide (502 mg, 12.5 mmol) and the reaction mixture stirred at
892 an ambient temperature for 16 h. Upon completion, the reaction mixture was acidified to pH
893 of 2-3 using 2 M aq. hydrochloric acid solution. The resulting precipitate was collected by
894 vacuum filtration and dried under reduced pressure to afford the desired acid as a white solid
895 (1.21 g, 84%). m.p. 250–252 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 8.10 (s, 1H, H-6), 8.05 (d,
896 J 8.2, 2H, H-3), 7.99 (d, J 8.1, 2H, H-2), 2.73 (s, 3H, CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ
897 167.5 (CO), 166.4 (C-8), 153.2 (C-5), 138.5 (c-4), 130.3 (C-2), 126.4 (C-3) , 116.5 (C-6), 19.4
898 (12-CH3). Quaternary 5-C not visible; IR (neat, v_{max}, cm^{-1}) 2826, 1667, 1608, 1573, 1421,
899 1290, 1169; 256 LCMS (LCQ) Rt = 0.6 min (method 1), m/z (ESI+) 220.2 $[M+H]^+$; HRMS m/z
900 (ESI): calcd. for $C_{11}H_9NO_2S$ $[M+H]^+$ 220.0427, found 220.0428.
901 *N-(2-methyl-5-sulfamoyl-phenyl)-4-(2-methylthiazol-4-yl)benzamide (**1**)*
902 To 4-(2-methylthiazol-4-yl)benzoic acid (120 mg, 0.55 mmol), HBTU [(2-(1H-benzotriazol-1-
903 yl)-1,1,3,3-tetramethyluronium hexafluorophosphate] (153 mg, 0.66 mmol), *N,N*-
904 diisopropylethylamine (191 µL, 1.09 mmol) in *N,N*-dimethylformamide (2 mL) was added 3-
905 amino-4-methylbenzenesulfonamide (102 mg, 0.55 mmol). The reaction mixture was stirred
906 at ambient temperature for 16 h. Upon completion, the solvent was removed under reduced
907 pressure. The crude product was taken up in ethyl acetate (5 mL), washed with saturated aq.
908 $NaHCO_3$ (4 mL), brine (4 mL), dried over $MgSO_4$, filtered and concentrated under reduced
909 pressure. The crude was purified by column chromatography (silica 12 g, 0 to 55% ethyl

910 acetate in petroleum ether) and further purified by column chromatography (amino silica 4 g,
911 0 to 5% methanol in dichloromethane) to yield the desired amide **1** as a white solid (15 mg,
912 7%). Rf 0.14 (petroleum ether:ethyl acetate 9:11); m.p. 261–263 °C; ^{1}H-NMR (500 MHz,
913 DMSO-d6) δ 10.11 (s, 1H, CONH), 8.15 – 8.09 (m, 3H, H-3, H-6), 8.06 (d, J 8.2, 2H, H-2),
914 7.87 (s, 1H, H-6'), 7.62 (d, J 8.0, 1H, H-4'), 7.48 (d, J 8.0, 1H, H9'), 7.34 (s, 2H, SO_2NH_2), 2.75
915 (s, 3H, 8-CH_3), 2.32 (s, 3H, 2'-CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.4 (CO), 165.5 (C-
916 8), 153.3 (C-5), 142.5 (C-5'), 138.3 (C-1), 137.6 (C-2'), 137.2 (C-2'), 133.6 (C-4), 131.2 (C-3'),
917 128.75 (C-2), 126.3 (C3), 124.1 (C-6'), 123.4 (C-4'), 116.1 (C-6), 19.4 (8-CH3), 18.44 (2'-
918 CH3); IR (neat, v_{max}, cm^{-1}) 3258, 2923, 1630, 1572, 255 1516, 1444, 1403, 1304, 1154; LCMS
919 (LCQ) Rt = 2.7 min (method 1), m/z (ESI+) 388.1 $[M+H]^+$; HRMS m/z (ESI): calcd. for
920 $C_{18}H_{17}N_3O_3S_2$ $[M+H]^+$ 388.6847, found 388.6850.
921
922 ***N*-(2,3-dimethyl-5-sulfamoyl-phenyl)-4-(2-methylthiazol-4-yl)benzamide (2)**
923 To 4-(2-methylthiazol-4-yl)benzoic acid (120 mg, 0.55 mmol), HBTU (153 mg, 0.66 mmol),
924 *N,N*-diisopropylethylamine (0.19 mL, 1.09 mmol) in *N,N*-dimethylformamide (3 mL) was added
925), 3-amino-4,5-dimethylbenzenesulfonamide (110 mg, 0.55 mmol). The reaction mixture was
926 stirred at ambient temperature for 16 h. Upon completion, the solvent was removed under
927 reduced pressure. The crude product was taken up in ethyl acetate (5 mL), washed with
928 saturated aq. $NaHCO_3$ (4 mL), brine (4 mL), dried over $MgSO_4$, filtered and concentrated under
929 reduced pressure. The crude was purified by column chromatography (silica 12 g, 0 to 60%
930 ethyl acetate in petroleum ether) to yield the desired amide **2** as a white solid (15 mg, 7%). Rf
931 0.12 (petroleum ether:ethyl acetate 1:1); m.p. 246–248 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ
932 10.17 (s, 1H, CONH), 8.13 – 8.08 (m, 3H, H-3, H-6), 8.05 (d, J 8.4, 2H, H-2), 7.64 (d, J 1.9,
933 1H, H-4'), 7.54 (d, J 1.9, 1H, H-3'), 7.28 (s, 2H, SO_2NH_2), 2.74 (s, 3H, 8-CH_3), 2.36 (s, 3H, 3'-
934 CH_3), 2.17 (s, 3H, 2'-CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.4 (CO), 165.7 (C-8), 153.3
935 (C-5), 141.6 (C-5'), 138.5 (C-1), 137.6 (ArC), 137.5 (ArC), 137.1 (ArC), 133.7 (C-4), 128.7 (C-
936 2), 126.3 (C-3), 124.6 (C-6'), 122.3 (C-4'), 116.1 (C-6), 20.7 (3'-CH3), 19.4 (8-CH3), 15.0 (2'-
937 CH3); IR (neat, v_{max}, cm^{-1}) 3258, 2923, 1630, 1572, 1516, 1444, 1403, 1304, 1154; LCMS
938 (LCQ) Rt = 2.0 min (method 1), m/z (ESI+) 402.1 $[M+H]^+$; HRMS m/z (ESI): calcd. for
939 $C_{19}H_{19}N_3O_3S_2$ $[M+H]^+$ 401.0868, found 401.0866
940
941 ***N*-(3-hydroxy-2-methyl-phenyl)-4-(2-methylthiazol-4-yl)benzamide (3)**
942 To 4-(2-methylthiazol-4-yl)benzoic acid (80 mg, 0.36 mmol) in dichloromethane (3 mL) was
943 added oxalyl chloride (37 µL, 0.44 mmol) in a dropwise manner followed by the addition of a
944 few drops of N,N-dimethylformamide (10 µL). The reaction mixture was stirred at ambient
945 temperature for 16 h. The crude acyl chloride was then added to a stirred mixture of 3-amino-
946 2-methylphenol (54 mg, 0.44 mmol), *N,N*-diisopropylethylamine (397 µL, 2.28 mmol) and

947 dichloromethane (1 mL) and stirred at ambient temperature for 2-16 h. The solvent was
948 removed under reduced pressure. The crude product was taken up in saturated aq. $NaHCO_3$
949 (5 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic components were
950 then washed with brine (10 mL), dried over $MgSO_4$, filtered and concentrated under reduced
951 pressure. The crude was purified by column chromatography (silica 12 g, 0 to 40% ethyl
952 acetate in petroleum ether) to yield the desired amide **3** as a white solid (13 mg, 10%).Rf 0.47
953 (petroleum ether:ethyl acetate 1:1); m.p. 238–240 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 9.84
954 (s, 1H, CONH), 9.35 (s, 1H, OH), 8.10 (s, 1H, H-6), 8.07 (d, J 8.2, 2H, H-3), 8.03 (d, J 8.3, 2H,
955 H-2), 7.00 (t, J 7.9, 1H, H-5'), 6.79 (d, J 7.8, 1H, H-6'), 6.73 (d, J 8.0, 1H, H-4'), 2.74 (s, 3H,
956 8-CH3), 2.03 (s, 3H, 2'-CH3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.4 (CONH), 165.3 (C-8),
957 156.2 (C-5), 153.4 (C-3'), 137.9 (ArC), 137.3 (ArC), 134.1 (ArC), 128.7 (H-2), 126.2 (H-3),
958 126.0 (ArC), 121.3 (ArC), 118.0 (C-6'), 116.0 (C-6), 112.9 (C-4'), 19.4, (C-8) 11.4 (C-2'); IR
959 (neat, v_{max}, cm^{-1}) 3291, 1640, 1607, 1499, 1466, 1307, 1174; LCMS (LCQ) Rt =2.5 min
960 (method 1), m/z (ESI+) 325.1 $[M+H]^+$; HRMS (ESI): calcd. for $C_{18}H_{16}NaN_2OS$ $[M+Na]^+$
961 347.0825, found 347.0827.
962
963 **Methyl 4-methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzoate (4)**
964 To 4-(2-methylthiazol-4-yl)benzoic acid (800 mg, 3.65 mmol) and methyl 3-amino-4-
965 methylbenzoate (52 µL, 9.12 mmol) in tetrahydrofuran (15 mL) was added phosphorus
966 trichloride (0.32 mL, 3.65 mmol). The reaction mixture was heated in a microwave reactor for
967 20 min at 150 °C. The crude product was taken up in saturated aq. $NaHCO_3$ (5 mL) and
968 extracted with ethyl acetate (3 × 10 mL). The combined organic components were then
969 washed with brine (10 mL), dried over $MgSO_4$, filtered and concentrated under reduced
970 pressure. The crude was purified by column chromatography (amino silica 12 g, 0 to 50%
971 ethyl acetate in petroleum ether) to yield the desired amide **4** as a white solid (920 mg, 65%).
972 Rf 0.23 (petroleum ether:ethyl acetate 3:1); m.p. 179–181 °C; ^{1}H-NMR (500 MHz, DMSO-d6)
973 δ 10.03 (s, 1H, CONH), 8.14 – 8.07 (m, 3H, H-3, H-6), 8.05 (d, J 8.3, 2H, H-2), 8.00 (d, J 1.8
974 , 1H, H-6'), 7.76 (dd, J 7.9, 1.8 , 1H, H-4'), 7.44 (d, J 7.9, 1H, H-9'), 3.85 (s, 3H, OCH3), 2.74
975 (s, 3H, 8-CH3), 2.33 (s, 3H, 2'-CH3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.4 (CO), 165.6 (C-
976 8), 153.3 (C-5), 140.0 (C-8'), 137.6 (C-1), 137.3 (C-7'), 133.7 (C-4), 131.3 (C-3'), 128.8 (C-2),
977 128.1 (C-6'), 127.5 (C-5'), 126.9 (C-4'), 126.3 (C-3), 116.1 (C-6), 52.5 (OCH3), 19.4 (8-CH3),
978 18.6 (2'-CH3). COO not visible; IR (neat, v_{max}, cm^{-1}) 3256, 1726, 1643, 1523, 1432, 1296,
979 1171, 1110; LCMS (LCQ) Rt = 2.6 min (method 1), m/z (ESI+) 367.0 $[M+H]^+$; HRMS m/z
980 (ESI): calcd. for $C_{20}H_{18}N_2O_3S$ $[M+Na]^+$ 389.0930, found 389.0931.
981
982 **4-Methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzamide (5)**
983 *Methyl 4-methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzoate*

To 4-(2-methylthiazol-4-yl)benzoic acid (800 mg, 3.65 mmol) and methyl 3-amino-4-methylbenzoate (52 µL, 9.12 mmol) in tetrahydrofuran (15 mL) was added phosphorus trichloride (0.32 mL, 3.65 mmol). The reaction mixture was heated in a microwave reactor for 20 min at 150 °C. The crude product was taken up in saturated aq. $NaHCO_3$ (5 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic components were then washed with brine (10 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude was purified by column chromatography (amino silica 12 g, 0 to 50% ethyl acetate in petroleum ether) to yield the desired amide as a white solid (920 mg, 65%). Rf 0.23 (petroleum ether:ethyl acetate 3:1); m.p. 179–181 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 10.03 (s, 1H, CONH), 8.14 – 8.07 (m, 3H, H-3, H-6), 8.05 (d, J 8.3, 2H, H-2), 8.00 (d, J 1.8 , 1H, H-6'), 7.76 (dd, J 7.9, 1.8 , 1H, H-4'), 7.44 (d, J 7.9, 1H, H-9'), 3.85 (s, 3H, OCH_3), 2.74 (s, 3H, 8-CH_3), 2.33 (s, 3H, 2'-CH3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 166.4 (CO), 165.6 (C-8), 153.3 (C-5), 140.0 (C-8'), 137.6 (C-1), 137.3 (C-7'), 133.7 (C-4), 131.3 (C-3'), 128.8 (C-2), 128.1 (C-6'), 127.5 (C-5'), 126.9 (C-4'), 126.3 (C-3), 116.1 (C-6), 52.5 (OCH_3), 19.4 (8-CH_3), 18.6 (2'-CH_3). COO not visible; IR (neat, v_{max}, cm^{-1}) 3256, 1726, 1643, 1523, 1432, 1296, 1171, 1110; LCMS (LCQ) Rt = 2.6 min (method 1), m/z (ESI+) 367.0 $[M+H]^+$; HRMS m/z (ESI): calcd. for $C_{20}H_{18}N_2O_3S$ $[M+Na]^+$ 389.0930, found 389.0931.

4-Methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzoic acid

To methyl 4-methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzoate (920 mg, 2.51 mmol) in methanol (24 mL) was added 2 M aq. sodium hydroxide (12.6 mL, 25.1 mmol) and the mixture stirred at an ambient temperature for 16 h. The reaction mixture was then heated for 2 h at 50 °C. The solvent was removed under reduced pressure. The crude was taken up in water and the mixture was acidified to a pH of 2–3 with 2 M aq. hydrochloric acid solution. The resulting precipitate was collected by vacuum filtration and dried under reduced pressure to afford the desired acid as a white solid (926 mg, 99%). m.p. 254–256 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 12.85 (s, 1H, COOH), 10.04 (s, 1H, CONH), 8.13 – 8.07 (m, 3H, H-3, H-6), 8.06 (d, J 8.3, 2H, H-2), 7.95 (d, J 1.7, 1H, H-6'), 7.74 (dd, J 7.9, 1.7, 1H, H-4'), 7.40 (d, J 7.9, 1H, H-3'), 2.74 (s, 3H, 8-CH_3), 2.32 (s, 3H, 2'-CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 167.4 (COOH), 166.4 (CONH), 165.6 (C-8), 153.3 (C-5), 139.5 (C-2'), 137.5 (C-1), 137.1 (C-1'), 133.8 (C-4), 131.1 (C-3'), 129.3 (C-5'), 128.7 (C-2), 127.8 (C-6'), 127.2 (C-4'), 126.3 (C-3), 116.01 (C-6), 19.4 (8-CH_3), 18.59 (2'-CH_3); IR (neat, v_{max}, cm^{-1}) 2923, 1679, 1638, 1512, 1492, 1414, 1390, 1251, 1178; LCMS (LCQ) Rt =3.0 min (method 1), m/z (ESI+) 353.0 $[M+H]^+$; HRMS (ESI): calcd. for $C_{19}H_{16}N_2NaO_3S$ $[M+Na]^+$ 375.0774, found 375.0774.

4-Methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzamide (5)

To 4- methyl-3-((4-(2-methylthiazol-4-yl)benzoyl)amino)benzoic acid (146) (100 mg, 0.28 mmol), EDCI (68 mg, 0.35 mmol), HOBt (54 mg, 0.35 mmol), *N,N*-diisopropylethylamine (100 µL, 0.57 mmol) in *N,N*-dimethylformamide (1 mL) was added 2 M ammonia in methanol (0.43 mL, 0.85 mmol). The reaction mixture was stirred at ambient temperature for 16 h. The solvent was removed under reduced pressure. The crude product was taken up in ethyl acetate (5 mL), washed with 1 M aq. hydrochloric acid solution. (4 mL), saturated aq. $NaHCO_3$ (4 mL), brine (4 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude was purified by column chromatography (amino silica 12 g, 0 to 5% methanol in dichloromethane) to yield the desired primary amide **5** as a white solid (39 mg, 37%)**.** Rf 0.17 (dichloromethane:methanol 19:1); m.p. 215–217 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 10.02 (s, 1H, CONH), 8.14 – 8.08 (m, 3H, H-3. H-6), 8.05 (d, J 8.5, 2H, H-2), 7.91 (s, 1H, 5'-CONHAB), 7.86 (d, J 1.8, 1H, H-6'), 7.70 (dd, J 7.9, 1.8, 1H, H-4'), 7.35 (d, J 7.9, 1H, H-3'), 7.27 (s, 1H, 5'-CONHAB), 2.74 (s, 3H, 8-CH_3), 2.28 (s, 3H, 2'-CH_3); ^{13}C-NMR (126 MHz, DMSO-d6) δ 167.9 (CO), 166.4 (CO), 165.5 (C-8), 153.3 (C-5), 138.0 (C-2'), 137.5 (C-1), 136.8 (C-1'), 133.8 (C-4), 132.8 (C-5'), 130.6 (C-3'), 128.7 (C-2), 126.6 (C-6'), 126.3 (C-3), 125.6 (C-4), 116.0 (C-6), 19.4 (8-CH_3), 18.37 (2'-CH_3); IR (neat, ν_{max}, cm^{-1}) 3108, 1713, 1668, 1608, 1571, 1501, 1436, 1410, 1278; LCMS (LCQ) Rt =1.7 min (method 1), m/z (ESI+) 352.0 $[M+H]^+$; HRMS (ESI): calcd. for $C_{19}H_{17}N_3NaO_2S$ $[M+Na]^+$ 374.0934, found 374.0933.

4-Methyl-3-((4-(2-methyloxazol-4-yl)benzoyl)amino)benzamide (6)

Methyl 4-(2-methyloxazol-4-yl)benzoate

Methyl 4-(2-bromoacetyl)benzoate (400 mg, 1.56 mmol) was stirred in neat acetamide (276 mg, 4.67 mmol) at 160 °C for 2 h. Water was added to the reaction mixture. The resulting precipitate was collected by vacuum filtration and dried under reduced pressure to afford the desired compound 191 as a brown solid (308 mg, 87%) which was carried forward without further purification. ^{1}H-NMR (500 MHz, DMSO-d6) δ 8.07 (s, 1H, H-6), 8.00 (d, J 8.2, 2H, H-3), 7.89 (d, J 8.1, 2H, H-2), 3.86 (s, 3H, OCH_3), 2.48 (s, 3H, 8-CH_3).

4-(2-Methyloxazol-4-yl)benzoic acid

To methyl 4-(2-methyloxazol-4-yl)benzoate (191) (279 mg, 1.2 mmol) in methanol (5 mL) and water (1.5 mL) was added sodium hydroxide (150 mg, 3.86 mmol) and the reaction mixture stirred at ambient temperature for 16 h. Upon completion, the reaction mixture was acidified to pH of 2-3 using 2 M aq. hydrochloric acid solution. The resulting precipitate was collected by vacuum filtration and dried under reduced pressure to afford the desired acid as a as a white solid (130 mg, 47%). ^{1}H-NMR (500 MHz, DMSO-d6) δ 8.56 (s, 1H, 11, H-6), 7.97 (d, J 8.0, 2H, H-3), 7.85 (d, J 8.0, 2H, H-2), 2.47 (s, 3H, 8- CH_3); LCMS (LCQ) Rt = 0.5 min (method 1), m/z (ESI+) 204.2 [M+H]+ .

4-Methyl-3-((4-(2-methyloxazol-4-yl)benzoyl)amino)benzamide (6)

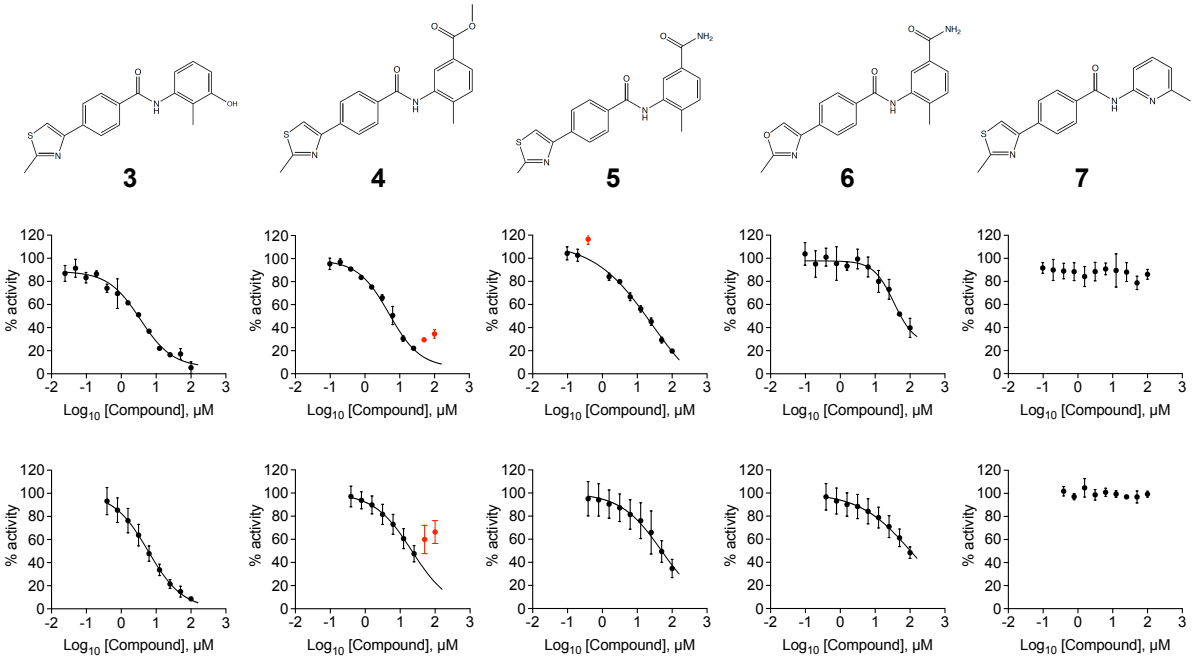
1057 To 4-(2-methylthiazol-4-yl)benzoic acid (110 mg, 0.54 mmol) in dichloromethane (3 mL) was
1058 added oxalyl chloride (0.05 mL, 0.65 mmol) in a dropwise manner followed by *N,N*-
1059 dimethylformamide (10 µL). The reaction mixture was stirred at ambient temperature for 16 h.
1060 The crude acyl chloride was then added to a stirred mixture of 3-amino-4-methylbenzamide
1061 (98 mg, 0.65 mmol) followed by the addition of *N,N*-diisopropylethylamine (0.47 mL, 2.71
1062 mmol). The reaction mixture was stirred at ambient temperature for 2-16 h. The solvent was
1063 removed under reduced pressure. The crude product was taken up in saturated aq. $NaHCO_3$
1064 (5 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic components were
1065 then washed with brine (10 mL), dried over $MgSO_4$, filtered and concentrated under reduced
1066 pressure. The crude was purified by column chromatography (silica 12 g, 0 to 5% methanol in
1067 dichloromethane) to yield the desired amide **6** as a white solid (55 mg, 29%). Rf 0.28
1068 (dichloromethane:methanol 19:1); m.p. 260–262 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 10.00
1069 (s, 1H, CONH), 8.61 (s, 1H, H-6), 8.05 (d, J 8.1, 2H, H-3), 7.91 (d, J 8.4, 3H, H-2, 5'-CONHAB),
1070 7.86 (s, 1H, H-6'), 7.70 (dd, J 7.9, 1.9, 1H, H-4'), 7.35 (d, J 7.9, 1H, H-3'), 7.27 (s, 1H, 5'-
1071 CONHAB) 2.28 (s, 3H, 2'-CH_3). 8-CH_3 under DMSO peak; ^{13}C-NMR (126 MHz, DMSO-d6) δ
1072 167.8 (CO), 165.4 (CO), 162.2 (C-8), 139.4 (C-5), 137.9 (C-2'), 136.8 (ArC), 136.3 (ArC),
1073 134.6 (C-1), 133.8 (C-4), 132.8 (C-5'), 130.6 (C-3'), 128.7 (C-2), 126.6 (C-6'), 125.5 (C-4'),
1074 125.33 (C-3), 18.4 (8-CH_3), 14.0 (2'-CH_3); IR (neat, , v_{max}, cm^{-1}) 3255, 1673, 1635, 1616, 1522,
1075 1489, 1386, 1276, 1214; LCMS (LCQ) Rt =0.8 min (method 1), m/z (ESI+) 336.1 $[M+H]^+$;
1076 HRMS (ESI): calcd. for $C_{19}H_{17}N_3NaO_3$ $[M+Na]^+$ 358.1162, found 358.1150.
1077
1078 ***N*-(6-methyl-2-pyridyl)-4-(2-methylthiazol-4-yl)benzamide (7)**
1079 To 4-(2-methylthiazol-4-yl)benzoic acid (80 mg, 0.36 mmol) in dichloromethane (3 mL) was
1080 added oxalyl chloride (37 µL, 0.44 mmol) in a dropwise manner followed by *N,N*-
1081 dimethylformamide (10 µL). The reaction mixture was stirred at ambient temperature for 16 h.
1082 The crude acyl chloride was then added to a stirred mixture of 2-amino-6-methylpyridine (40
1083 mg, 0.36 mmol) followed by the addition of *N,N*-diisopropylethylamine (397 µL, 2.28 mmol).
1084 The reaction mixture was stirred at ambient temperature for 2-16 h. The solvent was removed
1085 under reduced pressure. The crude product was taken up in saturated aq. $NaHCO_3$ (5 mL)
1086 and extracted with ethyl acetate (3 × 10 mL). The combined organic components were then
1087 washed with brine (10 mL), dried over $MgSO_4$, filtered and concentrated under reduced
1088 pressure. The crude was purified by column chromatography (silica 12 g, 0 to 45% ethyl
1089 acetate in petroleum ether) to yield the desired amide **7** as a colourless solid (54 mg, 45%).
1090 Rf 0.48 (petroleum ether:ethyl acetate 1:1); m.p. 195–197 °C; ^{1}H-NMR (500 MHz, DMSO-d6)
1091 10.69 (s, 1H), 8.13 (s, 1H), 8.10 (d, J 8.2, 2H), 8.06 (d, J 8.0, 2H), 8.02 (d, J 8.2, 1H), 7.73 (t,
1092 J 7.8, 1H), 7.03 (d, J 7.3, 1H), 2.74 (s, 3H), 2.46 (s, 3H); ^{13}C-NMR (126 MHz, DMSO-d6) 166.4,
1093 165.9, 157.0, 153.3, 152.0, 138.8, 137.6, 133.5, 129.1, 126.1, 119.5, 116.2, 112.1, 24.0, 19.4;

1094 IR (neat, v_{max}, cm^{-1}) 3314, 2923, 1610, 1539, 1522, 1290, 1169; LCMS (LCQ) Rt = 3.1 min
1095 (method 1), m/z (ESI+) 310.1 $[M+H]^+$; HRMS (ESI): calcd. for $C_{17}H_{16}N_3OS$ $[M+H]^+$ 310.1009,
1096 found 310.1012.
1097
1098 **1-(4-Fluoro-3-(trifluoromethyl)phenyl)-3-(5-(4-pyridyl)-1,3,4-thiadiazol-2-yl)urea**
1099 **(ML216)**[44]
1100 *Phenyl-5-(4-pyridyl)-1,3,4-thiadiazol-2-ylcarbamate*
1101 Sodium hydride (700 mg, 33.7 mmol) was slowly added to a suspension of 5-(4-pyridyl)-1,3,4-
1102 thiadiazol-2-yl-amine (2.00 g, 11.2 mmol) in tetrahydrofuran (40 mL) at 0 °C. The resulting
1103 reaction mixture was stirred at 0 °C for 2 h. Diphenyl carbonate (2.89 g, 13.5 mmol) was added
1104 and the reaction mixture was stirred at 0 °C for 30 min. The reaction mixture was warmed to
1105 ambient temperature and stirred overnight. Dichloromethane (40 mL) and brine (10 mL) was
1106 added to the reaction mixture and the solid precipitate was collected by filtration to yield the
1107 desired compound as a crystalline off-white solid (3.10 g, 93%). Rf 0.52
1108 (dichloromethane:methanol 19:1); m.p. 278–280 °C; ^{1}H-NMR (500 MHz, DMSO-d6) δ 8.58 (d,
1109 J 4.9, 2H, H-2), 7.70 (d, J 4.9, 2H, H-3), 7.33 (t, J 7.6, 2H, H-15), 7.11 (t, J 7.4, 1H, H-17), 7.07
1110 (d, J 7.7, 2H, H-16); ^{13}C-NMR (126 MHz DMSO-d6) δ = 174.7 (CO), 162.1 (ArC), 155.1 (ArC),
1111 153.7 (C-14), 150.7 (C-2), 140.1 (ArC), 129.2 (C-15), 124.0 (C-17), 122.4 (C-16), 120.2 (C-3);
1112 IR (neat, v_{max}, cm^{-1}) 3542, 3118, 2417, 1604, 1460, 1319, 1296, 1208, 1113; LCMS (LCQ) Rt
1113 = 1.8 min (method 2), m/z (ESI+) 299.03 $[M+H]^+$; HRMS (ESI): calcd. for $C_{14}H_{11}N_4O_2S$ $[M+H]^+$
1114 299.0598, found 299.0597.
1115 *1-(4-Fluoro-3-(trifluoromethyl)phenyl)-3-(5-(4-pyridyl)-1,3,4-thiadiazol-2-yl)urea (**ML216**)*
1116 To a suspension of phenyl -(5-(4-pyridyl)-1,3,4-thiadiazol-2-yl)carbamate (150 mg, 0.50 mmol)
1117 in toluene (5 mL) was added 4-fluoro-3-(trifluoromethyl)aniline (65 µL, 0.50 mmol). The
1118 reaction mixture was heated in a microwave reactor at 150 °C for 30 min and the resulting
1119 suspension was cooled to room temperature. The reaction mixture was concentrated under
1120 reduced pressure. The resulting solid 223 was triturated with dichloromethane (5 mL) and
1121 further triturated with 5% methanol in dichloromethane (3 mL) to yield the desired compound
1122 **ML216** as an orange solid (82 mg, 46%). ^{1}H-NMR (500 MHz, DMSO-d6) δ 11.61 (s, 1H, NH-
1123 urea), 9.52 (s, 1H, NH-urea), 8.73 (d, *J* 5.9, 2H, H-2), 8.06 (s, 1H, H15), 7.89 (d, *J* 5.3, 2H, H-
1124 3), 7.77 (s, 1H, H-19), 7.49 (t, *J* 9.7, 1H, H-18); LCMS (LCQ) Rt = 2.7 min (method 1), m/z
1125 (ESI+) 384.0 [M+H]+ . ^{1}H-NMR consistent with literature data[44].
1126

SUPPLEMENTARY TABLE 1. Statistics for data collection, phasing and refinement

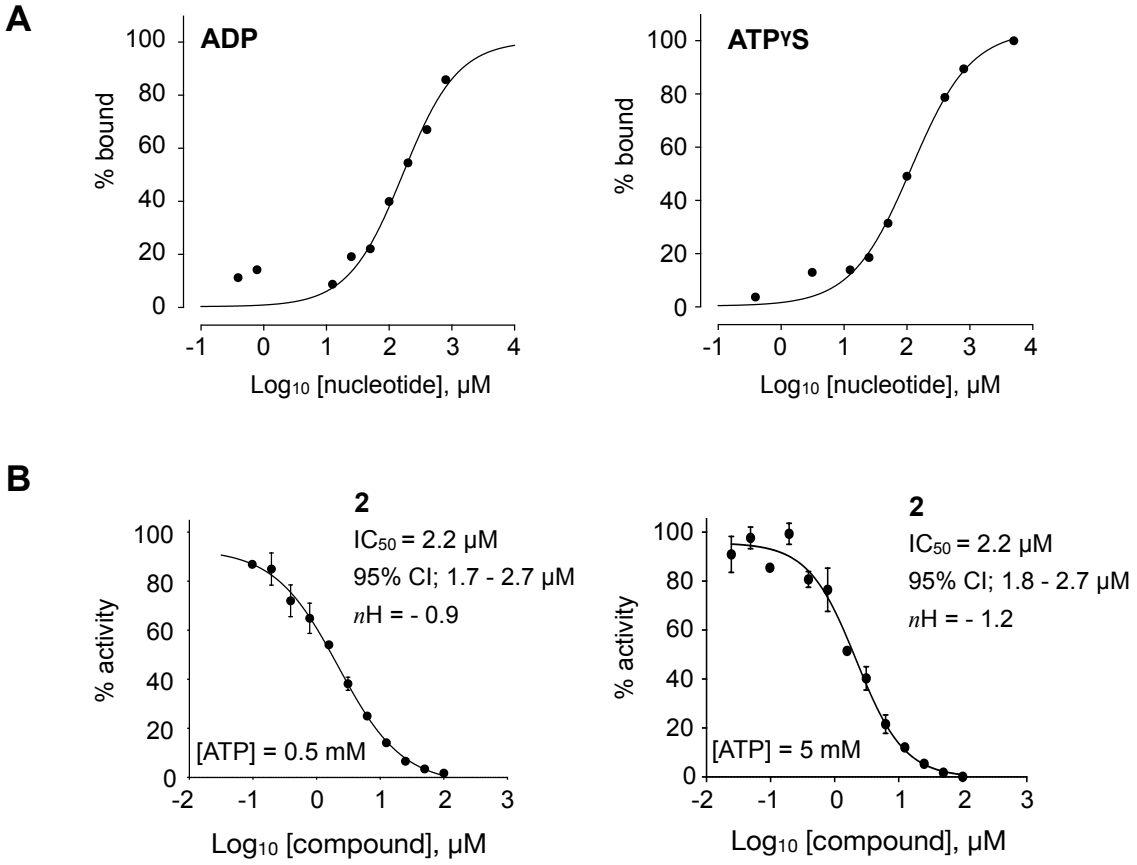
	BLM-HD$^{\Delta WHD}$ + ADP	Liganded-BLM-HD$^{\Delta WHD}$
Data collection		
Space group	$P2_1$	P1
Cell dimensions		
a, b, c (Å)	54.28, 107.69, 55.20	84.69, 111.60, 132.38
α, β, γ (º)	90.00, 109.31, 90.00	72.70, 80.13, 79.24
Wavelength	0.9780	0.9762
Resolution (Å)	51.23 – 1.53 (1.56– 1.53)	125.37 – 2.97 (3.08 – 2.97)
Mn I / σI	12.7 (1.2)	7.8 (1.4)
Mn I, $CC_{1/2}$	1.00 (0.61)	0.99 (0.57)
Completeness (%)	98.3 (90.3)	98.0 (94.5)
Redundancy	1.9 (1.7)	2.6 (2.7)
Refinement		
Resolution (Å)	51.23 – 1.53 (1.56– 1.53)	47.28-2.97 (3.07-2.96)
No. unique reflections	88464 (8096)	91661 (8892)
R_{work} / R_{free}	0.19 / 0.21	0.23 / 0.27
No. atoms		
Macromolecules	3870	24954
Ligands	65	427
Solvent	438	91
B-factors		
Wilson	32.58	77.77
ADP (mean)		
Macromolecules	31.00	95.79
Ligands	40.89	98.56
Solvent	45.69	53.65
R.m.s. deviations		
Bond lengths (Å)	0.014	0.006
Bond angles (º)	1.64	1.11
Molprobity		
All atom clashscore	2.44	7.42
Ramachandran		
Outliers	0.21 %	0.42 %
Allowed	1.65 %	3.72 %
Favoured	98.15 %	95.85 %

* values in parentheses are for the highest resolution shell


3
4
5
6
7
% activity
Log10 [Compound], μM

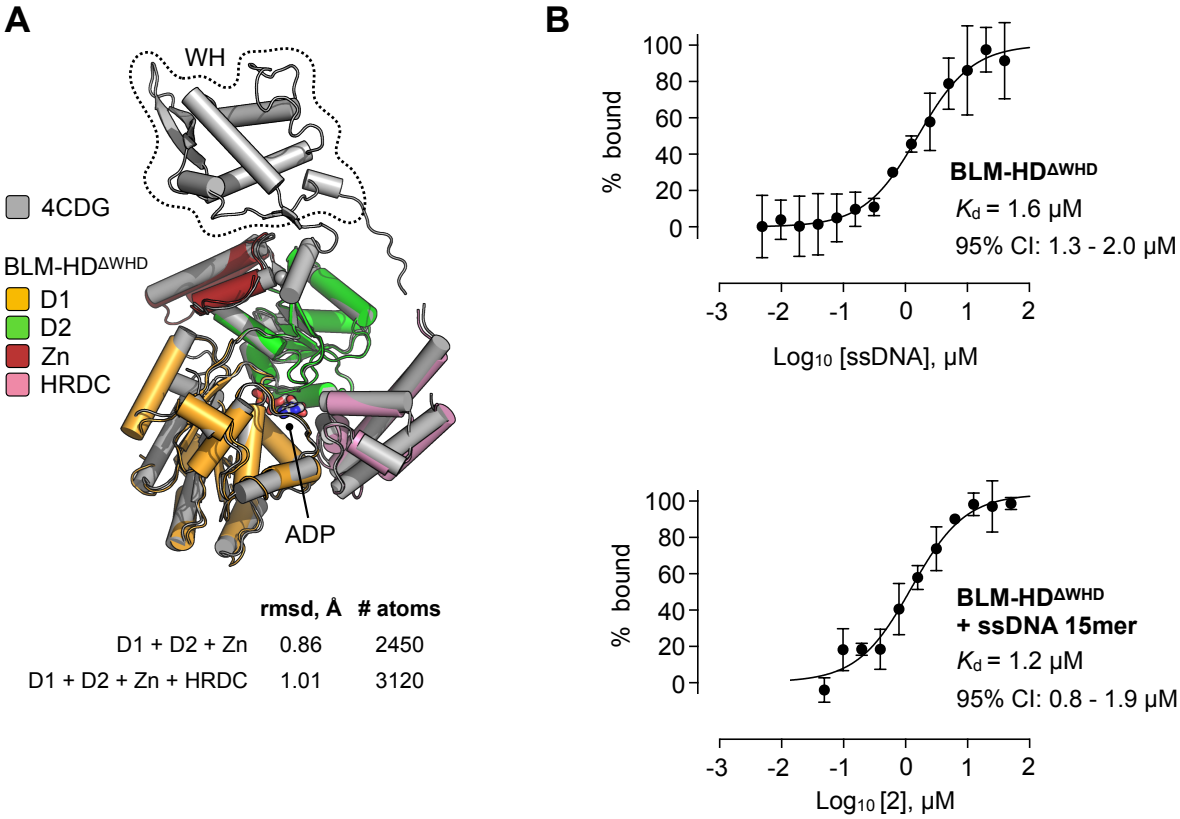
1127 **Figure S1**

1128 (Top) Chemical drawings for compounds **3** to **7.** (Middle) Dose response curves from
1129 fluorescence-based DNA unwinding assays with BLM-HD. Data points are the mean of three
1130 technical replicates with error bars representing 1 SD. **(**Bottom) Dose response curves from
1131 ATP turnover assays with BLM-HD. Data points represent the mean from at least two
1132 individual experiments, each containing 3 technical replicates. Error bars represent the
1133 standard error of the mean (SEM). Experimental data were fitted with a four parameter,
1134 log(inhibitor) vs. response model with variable slope. Calculated values for IC_{50}, Hill slope (*n*H)
1135 and 95% confidence intervals (95% CI) are reported in Table 1. Data points excluded from
1136 fitting are coloured red.


A
ADP
% bound
Log10 [nucleotide], μM
ATPγS
% bound
Log10 [nucleotide], μM
B
2
IC50 = 2.2 μM
95% CI; 1.7 - 2.7 μM
nH = - 0.9
[ATP] = 0.5 mM
% activity
Log10 [compound], μM
2
IC50 = 2.2 μM
95% CI; 1.8 - 2.7 μM
nH = - 1.2
[ATP] = 5 mM
% activity
Log10 [compound], μM

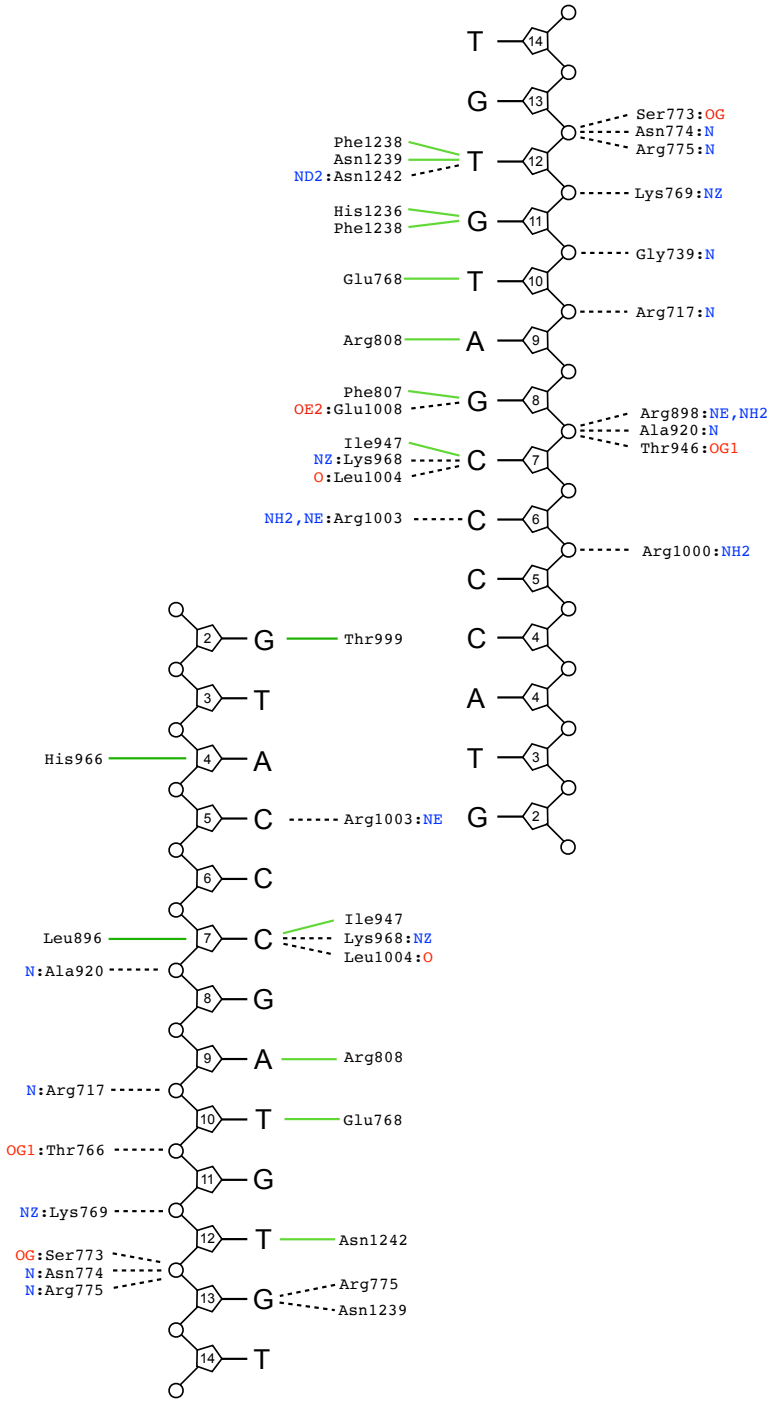
1137 **Figure S2**

1138 (A) Indicative binding isotherms for titrations of Mg-ADP and Mg-ATPγS into BLM-HD as
1139 measured by MST. Fitted lines are intended as visual aids only, as the data only represent
1140 values from a single preliminary experiment. (B) Dose response curves from fluorescence-
1141 based DNA unwinding assays with BLM-HD, carried out at two different ATP concentrations.
1142 Experimental data were fitted with a four parameter, log(inhibitor) vs. response model with
1143 variable slope. Calculated values for IC_{50}, Hill slope (*n*H) and 95% confidence intervals (95%
1144 CI) are given in each case. Data points represent the mean of 3 technical replicates, with
1145 error bars representing 1 SD.


A
WH
4CDG
BLM-HDΔWHD
D1
D2
Zn
HRDC
ADP
rmsd, Å # atoms
D1 + D2 + Zn 0.86 2450
D1 + D2 + Zn + HRDC 1.01 3120
B
% bound
100
80
60
40
20
0
BLM-HDΔWHD
Kd = 1.6 μM
95% CI: 1.3 - 2.0 μM
-3 -2 -1 0 1 2
Log10 [ssDNA], μM
% bound
BLM-HDΔWHD
+ ssDNA 15mer
Kd = 1.2 μM
95% CI: 0.8 - 1.9 μM
Log10 [2], μM

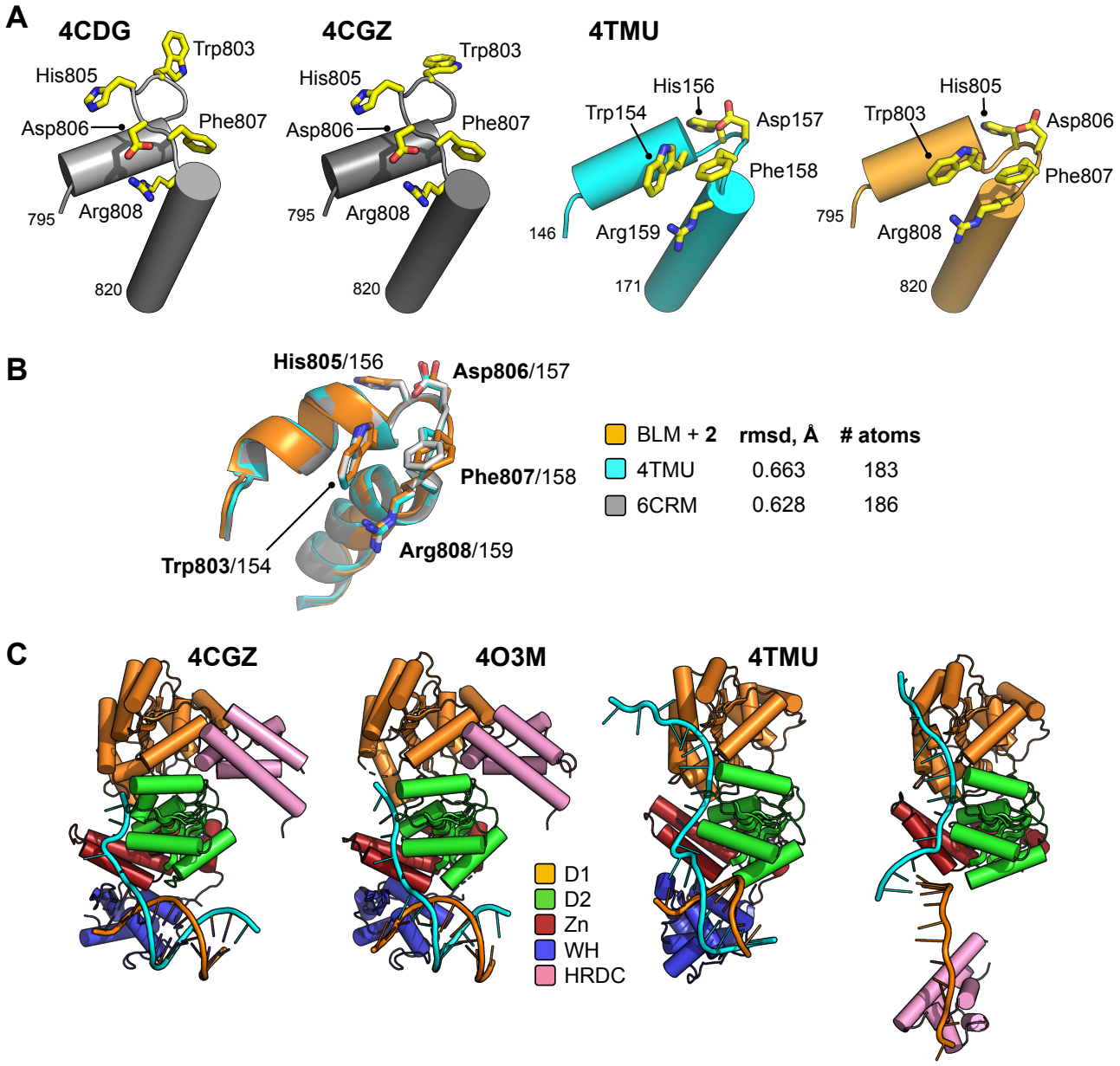
1146 **Figure S3**

1147 (A) Superposition of the structures of BLM-HD (PDB: 4CDG) and BLM-HD$^{\Delta WHD}$ using PyMOL
1148 [45]. Values for rmsd in Angstrom are shown, calculated over the indicated number of equivalent
1149 atoms. See associated key for details of colour scheme. (B) Isotherms for binding of BLM-
1150 HD$^{\Delta WHD}$ to (top) ssDNA-15mer and (bottom) compound **2** in the presence of ssDNA-15mer,
1151 as determined by MST. Values for K_d and 95% CI are given in each case. For all plots, data
1152 represent the mean of three technical replicates with error bars representing 1 SD.


T
G
T
G
T
A
G
C
C
C
C
A
T
G
Ser773:OG
Asn774:N
Arg775:N
Lys769:NZ
Gly739:N
Arg717:N
Arg898:NE,NH2
Ala920:N
Thr946:OG1
Arg1000:NH2
Phe1238
Asn1239
ND2:Asn1242
His1236
Phe1238
Glu768
Arg808
Phe807
OE2:Glu1008
Ile947
NZ:Lys968
O:Leu1004
NH2,NE:Arg1003
G
T
A
C
C
C
A
T
G
Thr999
His966
Arg1003:NE
Leu896
N:Ala920
Ile947
Lys968:NZ
Leu1004:O
Arg808
N:Arg717
Glu768
OG1:Thr766
NZ:Lys769
Asn1242
OG:Ser773
N:Asn774
N:Arg775
Arg775
Asn1239
T

1153 **Figure S4**

1154 Schematic summary of DNA-interactions made within the crystal lattice of the liganded
1155 complex (BLM-HD$^{\Delta WHD}$ + Mg-ADP + ssDNA-15mer + compound **2**). Interaction network
1156 determined through by the PDBsum Generate webserver[46]. Green solid lines represent
1157 hydrophobic / stacking interactions. Black dotted lines represent potential hydrogen bonds.


A
4CDG
Trp803
His805
Asp806
Phe807
795
Arg808
820
4CGZ
Trp803
His805
Asp806
Phe807
795
Arg808
820
4TMU
His156
Trp154
Asp157
Phe158
146
Arg159
171
His805
Trp803
Asp806
Phe807
795
Arg808
820
B
His805/156
Asp806/157
Phe807/158
Trp803/154
Arg808/159
BLM + 2 rmsd, Å # atoms
4TMU 0.663 183
6CRM 0.628 186
C
4CGZ
4O3M
4TMU
D1
D2
Zn
WH
HRDC

1158 **Figure S5**

1159 (A) Molecular secondary structure cartoons showing selected amino acid side chains of the
1160 aromatic rich loop (ARL) region in PDB entries 4CDG and 4CGZ (BLM-HD) and 4TMU
1161 (Cronobacter sakazakii RecQ) to that reported here for liganded-BLM-HDΔWHD. Amino acid
1162 numbering is provided for the start and end of the region compared, with the identity of key
1163 residues also provided. (B) Superposition of the ARL in liganded-BLM-HDΔWHD (coloured
1164 orange) with those found in PDB entries 4TMU (cyan) and 6CR[23] (grey); which both represent
1165 structures of the catalytic core of C. sakazakii RecQ in complex with different DNA substrates.
1166 (C) Molecular cartoon representations of the helicase catalytic cores reported in PDB entries
1167 4CGZ, 4O3M and 4TMU, highlighting their respective interactions with bound DNA substrates
1168 and comparing this to liganded-BLM-HD$^{\Delta WHD}$. Additional details for the colour scheme can be
1169 found in the associated key.